BÁLINT ÁGNES
TÁTRAI FERENC

GYAKORLATI STATISZTIKAI SZÁMÍTÁSOK

C 64

NOVOTRADE

BÁLINT ÁGNES — TÁTRAI FERENC

Statisztikai eljárások Commodore 64-esen

Novotrade, 1989

Lektorálta: Soltész Péter
Dr. Veress Gábor

A programok kidolgozásában részt vett:
Répás Zsuzsanna

Kiadásért felel RÉNYI GÁBOR, a Novotrade Rt. vezérigazgatója
Budapest, 1989

Műszaki szerkesztő: Erdősi Zoltán
Készült a Szegedi Nyomdában (13 A/5 ív)
Felelős vezető: SURÁNYI TIBOR igazgató

ISBN 963 585 003 4

Copyright © Dr. Bálint Ágnes, Dr. Tátrai Ferenc, 1989

TARTALOMJEGYZÉK

ELŐSZÓ

Egy hazai, statisztikával (is) foglalkozó tekintélyes professzorunk így kezdte első előadását:
„Jelenlegi ismereteink szerint a hazugság 3 kategóriája létezik:

— kis hazugság,
— nagy hazugság,
— statisztika."

Valóban, a statisztika tudományával akarva-akaratlanul sokan visszaéltek már. Mindezen visszaélések természetesen nem a statisztika eredendő bűnei: a statisztika „avatott" használatával döntéseink, helyzetünk és a változások trendjének megítélése megalapozottabb, biztosabb lehet.

A statisztika, a statisztikai módszerek ma már számítógépes kultúránk szerves részévé váltak.

A Novotrade Rt. gondozásában 1986-ban megjelent Voss professzor kitűnő könyve, a Bevezetés a statisztikai számításokba C 64-esen. Ez a könyv számos, igen értékes alapfokú statisztikai eljárás BASIC nyelvű programját tartalmazza. Jelen könyvünk megírásával nem a statisztikai témakörben megjelenő művek számát kívántuk eggyel szaporítani. Könyvünk alapjaiban eltér Voss könyvétől, de egyszersmind annak folytatásaként is felfogható.

Először is e könyvben a hangsúlyt nem a statisztikai módszerek programozásának részleteire fektettük, hanem a mélyebb statisztikai tartalomra. Könyvünk középfokúnak mondható. Nem az a célunk tehát, hogy a C 64-es kezelésében járatos Olvasót saját céljainak megfelelő, statisztikai programok készítésére tanítsuk meg, hanem kész programokat bocsátunk az Olvasó rendelkezésére, amellyel statisztikai adathalmaza feldolgozásával kapcsolatos kérdéseire választ tud kapni.

A könyv megírásakor olyan felhasználót tekintettünk „modellnek", aki nem akar programozási munkával bíbelődni, a munka érdemi részével szeretne törődni. Vannak már adatai (ezeket méréssel, kérdőíves feldolgozással vagy bármilyen más adatgyűjtési módszerrel szerezhette), és most minél kevesebb felesleges munkával szeretné az adataiban rejlő információkat megismerni, és ennek alapján esetleg döntéseket hozni. Természetesen nem sajnálja az időt a statisztikai tartalom megismerésére: nem akar téves döntéseket hozni. Milyen döntésekről van itt szó? A teljesség igénye nélkül felsorolunk néhány tipikus statisztikai döntési helyzetet.

I. Mezőgazdasági szakemberünk egy búzaszállítmány átvételéről vagy visszaküldéséről akar dönteni. Laboratóriumi mérésekkel, véletlen minták alapján ismeri a búza minőségi jellemzőinek átlagértékét és szórását, és főnökei meghatározták ezen jellemzők előírt értékeit. Az előírt és mért jellemzők természetesen eltérnek egymástól. Ha a szállítmányt átveszi, de a búza minősége rossz, megbüntetik. Ha nem veszi át,

de a szállítmány mégis jó volt, vállalatának kárt okozott: álltak a takarmánykeverő gépek, a szállító pereskedik stb. Hogyan döntsön?

II. Egy környezetvédelmi előadó szennyvízbírságot akar kiróni termelőszövetkezetünkre, mert a közeli folyó vizének foszfáttartalma meghaladja a megengedettet. Ezt természetesen nem folyamatos méréssel, hanem adott számú mintából állapította meg. Saját méréseink mást mutatnak. A mérési módszer szórása és a két mérés átlagának eltérése ismert. Reménykedhetünk-e benne, hogy az eltérés csak a véletlen műve volt, és új ellenőrzést kérve, megússza szövetkezetünk a büntetést, vagy jobb, ha nem szólunk egy szót sem, nehogy növekedjék a bírság?

III. Egy állami gazdaság két tehenészetében 15-15 tehén tejét elemezve a tej zsírtartalma kismértékben eltér. A mérési módszernek természetesen van szórása, a mért értékek is bizonyos szórást mutatnak. Hány tehén tejét kell megvizsgálnunk ahhoz, hogy az ismert korlátokkal rendelkező mérések alapján az egyik istállóban adott takarmányt jobbnak ítélhessük a másiknál?

IV. Színhelyünk egy épületkerámia-gyár. Tudjuk elméleti megfontolásokból, hogy a fagyálló padlólap égetés közben bekövetkező zsugorodása a felhasznált agyagkeverék számos tulajdonságától függ:

— az agyagásvány kalcium/alumínium arányától,
— az agyagkeverék nedvességtartalmától,
— az előszárítás hőmérsékletétől és időtartamától stb.

Az agyagkeverék ismert tulajdonságai alapján szeretnénk megállapitani a várható zsugorodást, hogy ennek függvényében tudjuk a formázó automatán a kipréselendő lapok méretét beállítani. Ehhez szükség van az említett változók és a zsugorodás közötti függvénykapcsolat megkeresésére és a függvény állandóinak kiszámítására. A függvény konstansai azonban függenek a későbbi égetés „szigorúságától", azaz hőmérsékletétől, idejétől stb. Ezért időről időre újra meg kell azokat határozni. Ez a feladat számítógépes paraméterbecslést igényel (a regressziós egyenlet paramétereinek becslését).

Könyvünk és a benne levő programcsomag ilyen és ehhez hasonló kérdésekre adja meg a választ. A programcsomag felépítése olyan, hogy a felhasználó a program által feltett kérdésekre adott válaszokkal ki tudja választani a számára megfelelő programot, sőt az eredményeket is értelmezni tudja.

A könyv fejezetei egymásra épülnek. Alkalmas középfokú statisztikai tanfolyamok segédkönyveként éppúgy, mint egyéni tanuláshoz, vagy a benne levő programok futtatásával konkrét problémák megoldására.

ÚTMUTATÓ KÖNYVÜNK HASZNÁLATÁHOZ

Könyvünk első fejezetében ismertetjük a statisztikai programcsomag felépítését és használatát. A további fejezetekben az egyes programok matematikai statisztikai tartalmával ismertetjük meg az Olvasót.

A különböző fejezetekből kiderül, hogy az egyes programok milyen kérdés, ill. probléma megválaszolására alkalmasak. Az elméleti ismereteken kívül a fejezetek mindegyike tartalmaz mintapéldát is, amelyeket az első fejezetben bemutatott programcsomag segítségével meg is oldottunk.

A programcsomag felépítését ismerő olvasó számára tehát világosak lesznek a különböző fejezetekben előforduló, sokszor ismétlődő felszólítások, mint pl.: „Válasszuk ki a FŐMENÜ-ből az ELOSZLÁSOK menüt!" vagy pl. „Hívjuk be az ELOSZLÁS MENÜBŐL a Poisson eloszlás programját!" stb.

Minden bemutatott példánál közöljük a program által számított eredményt, ill. eredményeket, és ahol szükséges az eredmények magyarázatát is megadjuk.

Hogyan kell tehát használni ezt a könyvet?

Egyéni tanulás, a statisztikai alapismeretek elsajátítása és gyakorlása a könyv segítségével

A továbbiakban feltételezzük, hogy az Olvasó ismeri a Commodore 64-es és a lemezmeghajtó egység használatát, legalábbis képes behívni a lemezről a programokat LOAD paranccsal és azokat el tudja indítani a RUN paranccsal. Ez esetben először célszerű a gép mellett ülve elolvasni a programcsomag ismertetését tartalmazó első fejezetet, és mindjárt be is hívni a főmenüt, majd az egyes almenüket. Az almenük utolsó ágán visszatérhetünk a FŐMENÜ-höz. Ezzel képet kaphatunk a feldolgozott, elméleti alapokkal is alátámasztott anyagról. Ne indítsuk el egyelőre egyik részprogramot sem, csupán azt nézzük át, hogy az egyes programok (menüágak) milyen részfeladatok megoldását kínálják.

Ezután olvassuk el az adatkezelést leíró fejezetet. Ezt a statisztikai anyagtól elkülönítve, a könyv utolsó fejezete tartalmazza. Megértése nem kíván statisztikai alapismereteket, de mivel a statisztikai adatfeldolgozás során általában nagyobb adathalmaz ismételt feldolgozásáról van szó, célszerű adatainkat egyszer s mindenkorra felvinni a mágneslemezre, és a továbbiakban onnan lekérdezni. Sokat segíthet a tanulásban, ha az Olvasó az egyes fejezetekben közölt futtatási eredményeket csak akkor nézi meg, amikor a mintapéldát már önállóan (természetesen nem kéziszámítással, hanem az idevonatkozó program lefuttatásával) megoldotta. A kapott és az általunk közölt eredmények összehasonlítása hasznos tanulságokkal járhat.

A könyv használata statisztikai tanfolyamok oktatási segédleteként

Saját oktatási tapasztalataink azt mutatták, hogy a „száraz" statisztikai ismeretek oktatásában a példák megoldásának nagy meggyőző ereje van. A legtöbb statisztikai példa megoldása azonban egyszerű, ám sokszor ismétlendő számtani műveletek elvégzését igényli. Ez az előadásokat kellemetlen módon széttördeli, elvész az eredeti, megértetni kívánt gondolatmenet. Könyvünkben ezt úgy próbáltuk áthidalni, hogy az előadandó elméleti anyagot rögtön követi a példa megoldása. Az eredmény a demonstrációs eszközként használt számítógéppel a képernyőre másodpercek töredéke alatt kiírható, sőt a sornyomtatón kiírt eredményekből minden hallgatónak adhatunk. Ezt a demonstrációt elősegíti, hogy az elméleti ismertetőben és a programban ugyanazokat a szakkifejezéseket használjuk, így nem kell „lefordítani" a számítógép által kiírt eredményeket.

A könyv használata statisztikai számítások megoldásában

Tételezzük most fel, hogy az Olvasó tisztában van a könyvben feldolgozott statisztikai fogalmakkal, és konkrét, napi problémáiban akarja a könyvet és a programot felhasználni. Ez esetben azt javasoljuk, hogy az egész programot ismertető első fejezet és az adatkezelést ismertető fejezet elolvasása után vegye fel adatait mágneslemezre, és keresse meg azt a menüágat, amely feladatai megoldására alkalmazható. Ha a módszer lényegével tisztában vagyunk, a megfelelő program kiválasztása, futtatása és az eredmények értékelése nem okozhat gondot. Mágneslemezen tárolva az adatokat egyszerűen válogathatunk, módosíthatunk, ill. bővíthetünk stb. adatállományunkban.

Két jótanács: A programok legelső futtatása előtt írassuk ki a programismertetőt a képernyőre. Adattárolásra külön mágneslemezt használjunk.

A programcsomag továbbfejlesztése

A könyv minden fejezete tartalmazza a vonatkozó program ismertetését, a változók listáját és a programlista „izgalmas" részeit. Ezek alapos átnézése lehetőséget ad a program bővítésére, hiszen minden speciális feladat megoldására programunk nincs felkészítve, nem is lehet. Javasoljuk azonban, hogy programmódosítással csak az próbálkozzon, aki járatos a BASIC programozásban, és a statisztikai módszerek használatában is.

Könyvünk használatához sok sikert kívánunk.

A szerzők

1. A PROGRAMCSOMAG FELÉPÍTÉSE ÉS MŰKÖDÉSE

A programokat BASIC nyelven, Commodore 64-es típusú számítógépre írtuk. A könnyebb használhatóság és együttes kezelhetőség érdekében a programokat programcsomaggá szerveztük.

A programcsomag összeállításánál törekedtünk az interaktivitásra, így a programok futtatása azok számára sem okoz gondot, akik kevéssé járatosak a statisztikában és a programozásban. Valamennyi programot párbeszédes üzemmódban készítettük, sőt, néhol külön interaktív programok segítik a megfelelő programok kiválasztását. Ilyen például az INTERAKTÍV PROGRAM A MEGFELELŐ STATISZTIKAI PRÓBA KIVÁLASZTÁSÁRA című, amellyel az is könnyen eldöntheti, melyik statisztikai próbát kell alkalmazni az adott esetben, aki nem ismeri ezeket a próbákat.

A felhasználónak valamennyi programban kérdésekre kell válaszolnia, így viheti be az adatokat, választhatja ki a megfelelő menüágat vagy írathatja ki az eredményt.

A felhasználónak csak a FŐMENÜ című programot kell behívnia és elindítania, ezután maga a FŐMENÜ program hívja be azt a programot, amelyet kiválasztottunk, majd a program lefutása után visszatér a FŐMENÜ-re. Ha behívjuk és elindítjuk a FŐMENÜ című programot, akkor a következő felirat jelenik meg a képernyőn:

FŐMENÜ
STATISZTIKAI PROGRAMCSOMAGHOZ

AZ ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:

ADATFILE MŰVELETEK	(1)
ALAPVETŐ STATISZTIKAI JELLEMZŐK	(2)
ELOSZLÁSOK	(3)
STATISZTIKAI PRÓBÁK	(4)
SZÓRÁSELEMZÉS	(5)
REGRESSZIÓSZÁMÍTÁS	(6)
VÉGE	(7)

Vegyük sorra, mi történik, ha az egyes menüágakat választjuk! Bármely választás esetén, a program először is azt kéri, hogy tegyük be a lemez megfelelő oldalát, a programcsomag ugyanis két lemezoldalon található. Ha az ADATFILE MŰVELETEK-et választottuk, a program behívja az adatfeldolgozó programot.

Mivel a statisztikai programokkal általában igen sok adatot dolgozunk fel egyszerre, célszerű azokat háttértárolón rögzíteni és előkészíteni. Az adatfile-okat a programcsomag valamennyi programja használja, ill. használhatja, fontosságára való tekintettel az ADATFILE MŰVELETEK programot a fejezet végén külön

pontban a többinél részletesebben ismertetjük. Ha az ALAPVETŐ STATISZTIKAI JELLEMZŐK-et választottuk a következőket láthatjuk a képernyőn:

ALAPVETŐ STATISZTIKAI JELLEMZŐK
AZ ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:
ÁTLAG-, MEDIÁN- ÉS SZÓRÁSSZÁMÍTÁS (1)
MINTAFELDOLG. OSZTÁLYBA SOROLÁSSAL (2)
VISSZA A FŐMENÜHÖZ (3)

(Az osztályozás a könyvben szűkebb értelemben a minta elemeinek számszerű tulajdonságok szerinti csoportosítását jelenti. A mennyiségi osztályok alsó és felső határértékkel rendelkeznek, a kettő különbsége az osztályköz (osztásköz).)

Hivassuk be az átlag-, medián- és szórásszámító programot! Először a program címe jelenik meg, majd kívánság szerint kiírathatjuk az ismertetést. A program ezután az adatokat kéri, és azokat igény szerint lemezről (ha előzőleg rögzítettük) vagy billentyűzetről olvassa be. Ha ez megtörtént, a program kiszámítja a számtani, mértani, harmonikus, négyzetes és logaritmikus átlagot, a mediánt, a korrigált tapasztalati szórást és a szórásnégyzetet. Az eredmények kiírása után újabb adatokkal ismételhetjük meg ugyanezt vagy visszatérhetünk a FŐMENÜ-höz.

Ez a program azért nagyon hasznos számunkra, mert a statisztikai próbákat végző programokban lehetőség van arra, hogy ne a teljes adatsort, hanem csak a már előre kiszámított átlagot és szórást vigyük be. Ez nagymértékben megkönnyíti és meggyorsítja munkánkat és csökkenti a helyfoglalást a belső tárban.

Most hívassuk be a MINTAFELDOLGOZÁS OSZTÁLYBA SOROLÁSSAL című programot! A cím és az ismertetés kívánság szerinti kiíratása után az adatokat kéri a program, amelyeket lemezről is, billentyűzetről is bevihetünk. Ezután nagyság szerint sorba rendezi az adatokat és kiírja, melyik a legkisebb és a legnagyobb elem, megkérdezi, hány osztásközbe kívánjuk sorolni az adatokat, majd kéri az osztásközök felső határát. Az osztályba sorolás után lehetőség van a javításra, majd a menü jelenik meg a képernyőn, ahol kérhetjük a gyakorisági táblázat kiírását, a módusz kiszámítását, a tapasztalati eloszlásfüggvény, ill. a hisztogram felrajzolását vagy kiléphetünk a programból. Bármely menüágat választottuk, utána ehhez a menühöz tér vissza a program.

Ha a FŐMENÜ-ben az ELOSZLÁSOK című menüágat választottuk, a következőket láthatjuk a képernyőn:

ELOSZLÁSOK

AZ ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:

DISZKRÉT ELOSZLÁSOK:
BINOMIÁLIS ELOSZLÁS (1)
POISSON ELOSZLÁS (2)

FOLYTONOS ELOSZLÁSOK:
NORMÁLIS ELOSZLÁS (3)
EXPONENCIÁLIS ELOSZLÁS (4)
KHI-NÉGYZET ELOSZLÁS (5)
STUDENT ELOSZLÁS (6)
FISCHER ELOSZLÁS (7)

ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA (8)
VISSZA A FŐMENÜHÖZ (9)

Itt az ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA kivételével valamennyi program azonos módon épül fel, ezért ezeket együtt ismertetjük. A programok elején kérhetjük az ismertetés kiírását, majd meg kell adni az eloszlás paramétereit és a valószínűségi változó értékét (ezek helyességét a program ellenőrzi). Ezután kiszámítja az ehhez tartozó valószínűséget (normális eloszlás esetén az eloszlás- és a sűrűségfüggvény értékét), majd kiírja a képernyőre. Lehetőség van arra, hogy más értékekkel megismételjük a számolást.

Ha az ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSÁ-t választottuk, az előbbi diszkrét eloszlások hisztogramjait, ill. a folytonos eloszlások eloszlás- és sűrűségfüggvényeit rajzolhatjuk ki a képernyőre vagy a nyomtatóra. A program alapján a menüből kell kiválasztani az eloszlást, majd a felrajzolás után ide térünk vissza, hogy újabb eloszlásfüggvény felrajzolását kérjük, vagy kilépjünk a programból.

Ha a FŐMENÜ-ben a STATISZTIKAI PRÓBÁK című menüágat választottuk, a következő jelenik meg a képernyőn:

STATISZTIKAI PRÓBÁK

AZ ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:

A MEGFELELŐ PRÓBA INTERAKTÍV KIVÁLASZTÁSA	(1)
U STATISZTIKA HASZNÁLATA	(2)
T STATISZTIKA HASZNÁLATA	(3)
F-PRÓBA ÉS BARTLETT-PRÓBA	(4)
FÜGGETLENSÉGVIZSGÁLAT KHI-NÉGYZET PRÓBÁVAL	(5)
NORMALITÁS VIZSGÁLAT KHI-NÉGYZET PRÓBÁVAL	(6)
VISSZA A FŐMENÜHÖZ	(7)

A MEGFELELŐ PRÓBA INTERAKTÍV KIVÁLASZTÁSA című programot már az előbb említettük. Ez a program kérdések feltevésével segíti a problémánkra választ adó statisztikai próba kiválasztását, és be is hívja a kiválasztott programot.

Az U STATISZTIKA HASZNÁLATA című program a várható értékekre vonatkozó hipotézis vizsgálatát végzi el, amennyiben ismerjük a szórást. Először a menü jelenik meg, ahol választhatunk, hogy egy- vagy kétmintás próbát akarunk-e végezni, az ismertetést szeretnénk-e elolvasni vagy ki szeretnénk lépni a programból. Ezután az adatokat kéri a program, amelyeket természetesen itt is akár lemezről, akár billentyűzetről is bevihetünk. Ha billentyűzetről visszük be az adatokat választhatunk, hogy a teljes adatsor(oka)t akarjuk-e beolvasni vagy a már előre kiszámított átlago(ka)t.

Ezután következnek a számítások. Ha az egymintás u-próbát választottuk, megjelenik az egymintás u-próba almenüje, ahol a következő lehetőségeink vannak: az u-érték kiszámítása, a megbízhatósági intervallum kiszámítása, a mérések számának meghatározása, visszatérés a program főmenüjéhez vagy kilépés a programból (visszatérés a FŐMENÜ programhoz). A számítások elvégzése után mind az egymintás, mind a kétmintás változatnál a program megkérdezi, hogy egyoldali vagy kétoldali hipotézisről van-e szó, és hogy hány százalékos megbízhatósági szintet választunk, majd kiírja az eredményt. Ezután visszatérünk a menühöz (ha egymintás próbát számoltunk, akkor az egymintás próba almenüjéhez, ha kétmintás próbát számoltunk, akkor a program főmenüjéhez).

A T STATISZTIKA HASZNÁLATA című program szintén a várható értékekre vonatkozó hipotézist vizsgálja abban az esetben, ha nem ismerjük az elméleti szórásokat. A program hasonlóképpen épül fel, mint az EGY- ÉS KÉTMINTÁS U-PRÓBA. Az elsőként megjelenő menüben a következő lehetőségek közül választhatunk: egymintás t-próba, kétmintás t-próba, Welch-próba, az ismertetés kiíratása, kilépés a programból. Az adatbevitelre ugyanazok a lehetőségek, mint az u-próbánál. Az egymintás t-próba almenüje megegyezik az egymintás u-próbáéval. Ez a program is megkérdezi, hogy egyoldali vagy kétoldali hipotézisről van-e szó, és hogy hány százalékos megbízhatósági szintet választunk. Az eredmény kiírásának módja, majd a további lépések megegyeznek az előző programnál leírtakkal.

Az F-PRÓBA ÉS BARTLETT-PRÓBA című program a szórások egyezését vizsgálja, két és több változó esetén. Először szintén a menü jelenik meg, ahol választhatunk, hogy F-próbát vagy Bartlett-próbát akarunk-e végezni, az ismertetést akarjuk-e kiíratni vagy ki szeretnénk lépni a programból. Ezután az adatokat kell bevinni, a teljes adatsort vagy a már kiszámított szórásokat. A program nemcsak az F-értéket, ill. a χ^2 értékét számítja ki, hanem az azokhoz tartozó szignifikancia szintet, ill. valószínűséget is. Az eredmények kiírása után a menüre térünk vissza.

FÜGGETLENSÉGVIZSGÁLAT KHI-NÉGYZET PRÓBÁVAL olyan vizsgálat, amelyben arra a kérdésre keresünk választ, hogy két vagy több esemény valószínűsége megegyezik-e. A program legelején kívánság szerint kiírathatjuk az ismertetést, majd az adatokat kell bevinnünk. Az adatokat ebben az esetben az ún. kontingencia-táblázat tartalmazza. Ezután következnek a számítások, az eredmények kiíratása, majd újabb számítást végezhetünk vagy visszatérhetünk a FŐMENÜ-re.

Ha a NORMALITÁSVIZSGÁLAT KHI-NÉGYZET PRÓBÁVAL című programot választottuk, arra a kérdésre kaphatunk választ, hogy egy adott adatsor normális eloszlású-e. Az esetleges ismertetés kérése után következik az adatok bevitele. Lehetőségünk van a teljes adatsor előkészítés nélküli bevitelére is. Ebben az esetben a program nagyság szerint sorba rendezi az adatokat, kiírja melyik a legkisebb és legnagyobb elem, majd megkérdezi, hány osztályközbe akarjuk sorolni az adatokat, és kéri az osztályközök felső határait.

Lehetőség van arra is, hogy már osztályba sorolt adatokat vigyünk be, de akkor is meg kell adnunk az osztályközök felső határait és az egyes osztályközökhöz tartozó gyakoriságokat, valamint az átlagot és a szórást. Ezután következnek a számítások, majd az eredmény kiíratása. Nemcsak a χ^2 értékét írja ki a program, hanem a hozzá tartozó valószínűséget is. Lehetőség van újabb normalitásvizsgálat elvégzésére vagy visszatérhetünk a FŐMENÜ programra. Ha a FŐMENÜ-ben a SZÓRÁSELEMZÉS című menüágat választottuk, a következő jelenik meg a képernyőn:

SZÓRÁSELEMZÉS
Az ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:
INTERAKTÍV PROGRAM A MEGFELELŐ
  ALPROGRAM KIVÁLASZTÁSÁRA (1)
SZÓRÁSELEMZÉS EGYSZERES OSZTÁLYOZÁSSAL (2)
SZÓRÁSELEMZÉS KÉTSZERES OSZTÁLYOZÁSSAL (3)
VISSZA A FŐMENÜHÖZ (4)

Az INTERAKTÍV PROGRAM A MEGFELELŐ ALPROGRAM KIVÁLASZTÁSÁRA című program rövid magyarázattal segítséget nyújt a megfelelő program kiválasztására, és be is hívja azt.

A SZÓRÁSELEMZÉS EGYSZERES OSZTÁLYOZÁSSAL és SZÓRÁSELEM-

ZÉS KÉTSZERES OSZTÁLYOZÁSSAL című programok felépítése gyakorlatilag megegyezik. Mindkét programban először ismertetést kérhetünk, majd az adatokat kell bevinnünk, amit lemezről is, billentyűzetről is megtehetünk. Ezután következnek a számítások, majd az eredmény kiírása. A program kiírja a szórásfelbontó táblázatot és az F-értéket is, majd újabb elemzést végezhetünk vagy visszatérhetünk a FŐMENÜ programra. Ha a FŐMENÜ-ben a REGRESSZIÓSZÁMÍTÁS című menüágat választottuk, a következőket láthatjuk a képernyőn:

REGRESSZIÓSZÁMÍTÁS

AZ ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:
INTERAKTÍV PROGRAM A MEGFELELŐ
  ALPROGRAM KIVÁLASZTÁSÁRA (1)
  EGYVÁLTOZÓS REGRESSZIÓ:
  LINEÁRIS (2)
  EXPONENCIÁLIS (3)
  GEOMETRIAI (4)
  LOGARITMIKUS (5)
  ANTOINE-TÍPUSÚ (6)
  N-ED RENDŰ HATVÁNYSORRAL
    VALÓ KÖZELÍTÉS (7)
  TÖBBVÁLTOZÓS LINEÁRIS:
    A VÁLTOZÓK TRANSZFORMÁCIÓJA
    NÉLKÜL (8)
    A VÁLTOZÓK TRANSZFORMÁCIÓ-
    JÁVAL (9)
  KOVARIANCIAMÁTRIX SZÁMÍTÁSA (A)
  VISSZA A FŐMENÜHÖZ (B)

Az INTERAKTÍV PROGRAM A MEGFELELŐ ALPROGRAM KIVÁLASZTÁSÁRA című program itt is kérdések feltevésével segíti a megfelelő program kiválasztását, majd be is hívja azt.

Valamennyi egyváltozós regressziós, és a többváltozós lineáris regressziót végző program is azonos módon épül fel. Először az ismertetést írathatjuk ki, majd az adatokat kell bevinnünk, amit lemezről is, billentyűzetről is megtehetünk. Az adatbevitel után ellenőrzésre kiírathatjuk az adatokat a képernyőre. Ezután a program elvégzi a számításokat, majd kiírja a regressziós modell együtthatóit.

Kívánság szerint a regressziós táblázatot is kiírathatjuk, majd kérhetjük a mérési pontok mért és számított értékeinek listáját. (A regressziós táblázat a regressziószámításnál fellépő négyzetösszeg felbontását, az illesztett függvény illeszkedésének jellemzésére szolgáló T értéket, a korrelációs koefficiens számított értékét tartalmazza.)

Ezután az adott intervallumon belül a független változó bármely értékéhez kiszámíthatjuk a függő változót az illesztett görbe alapján. Végül újabb számítást kérhetünk vagy visszatérhetünk a FŐMENÜ-re.

Ha a KOVARIANCIAMÁTRIX SZÁMÍTÁSA című programot választottuk, először az ismertetés kiírását kérhetjük, majd az adatokat kell bevinnünk, lemezről vagy billentyűzetről. Az adatok helyességét ellenőrizhetjük a képernyőre való kiíratással. Ezután következnek a számítások, majd a program kiírja az eredményeket, végül újabb számítást kérhetünk vagy visszatérhetünk a FŐMENÜ-re. Természetesen valamennyi programban lehetőség van arra, hogy az eredményeket ne a képernyőre, hanem a nyomtatóra írassuk ki.

2. VALÓSZÍNŰSÉGSZÁMÍTÁSI ALAPFOGALMAK

2.1. A véletlen fogalma

Ha azt mondjuk, hogy egy jelenség véletlenszerű, ezen nem azt értjük, hogy annak, ami végbemegy, nincs oka. Oka mindennek van, csupán arról lehet szó, hogy ez az ok számunkra részben ismeretlen. Minden jelenséget az okok bizonyos rendszere hoz létre, s ha ezeknek csak egy részét vesszük tekintetbe, a jelenség véletlenszerű.

A mezőgazdaságban, a gazdasági elemző munkában, környezetvédelemben stb. igen sok véletlenszerű jelenséggel találkozunk. Ez indokolja a fenti tudományág alkalmazását mindennapi, gyakorlati munkánk tervezésénél és értékelésénél.

Véletlen kísérlet, véletlen esemény

Minden valószínűségi probléma valamilyen véletlen kísérlettel kapcsolatos. Véletlen kísérleten az olyan kísérletet értjük, amelynek kimenetelét az általunk figyelembe vett kísérleti körülmények nem határozzák meg egyértelműen. Megállapodás szerint kísérletnek nem csupán egy jelenség mesterséges előállítását nevezzük, hanem általánosságban egy jelenség megfigyelését függetlenül attól, hogy azt mi vagy rajtunk kívülálló okok hozták létre. Véletlen kísérlet pl.: egy folyó vízállásának megfigyelése egy meghatározott helyen és időpontban; az átlagos napi középhőmérséklet megfigyelése egy meghatározott napon stb.

Minden véletlen kísérletnek több, esetleg végtelen sok lehetséges kimenetele van. Az összes lehetséges kimenetelt együttesen *eseménytérnek* nevezzük (Ω). Dobókockával dobva nyilván csak a következő esetek lehetségesek: a dobás eredménye vagy 1, vagy 2, vagy 3, vagy 4, vagy 5, vagy 6. Ezek az eredmények azonban csoportosíthatók. Pl.: vannak közöttük páros számok (2, 4, 6), páratlan számok (1, 3, 5), vannak 3-nál kisebbek (1, 2) stb. Ezeket a csoportokat az osztályozás alapjául választott közös tulajdonságuk definiálja.

Az eseménytérnek valamilyen közös tulajdonság alapján kiválasztott pontjait együttesen *eseménynek* nevezzük. Esemény például az, hogy a dobás eredménye páros, ill. páratlan. Vizsgálhatjuk azt az eseményt, hogy egy folyó vízállása egy adott helyen nagyobb-e, mint 600 cm.

Gyakoriság, relatív gyakoriság, valószínűség

Ha egy véletlen kísérletet egyszer elvégzünk, akkor lehetséges kimenetelére semmilyen bizonyosságunk nincs, azaz semmilyen alapunk nincs arra, hogy a kísérlet egy jövőbeni kimenetelét megjósoljuk. A kísérletet többször megismételve fel lehet fedezni a véletlen események bekövetkezésében bizonyos törvényszerűségeket. Ez abban áll, hogy ha figyelembe vesszük az egyes események *gyakoriságát,* azaz azt,

hogy az illető esemény a kísérletsorozatban hányszor fordult elő, akkor az egyes események gyakoriságainak hányadosa bizonyos stabilitást mutat. Ez a stabilitás a következőképpen értendő.

Jelölje A az egyik eseményt, B a másik eseményt, k_A az A esemény gyakoriságát, k_B a B esemény gyakoriságát. Ekkor a k_A/k_B hányados viszonylag kis ingadozást mutat. Ez azt a feltételezést sugallja, hogy a k_A/k_B hányados a kísérletek számának minden határon túli növelésekor egy meghatározott számértékhez tart.

A valóságban természetesen nem lehet minden határon túl növelni a kísérletek számát. Ezért megelégszünk azzal, hogy nagy számú kísérlet esetén a k_A/k_B hányados gyakorlatilag állandó.

Ezen az alapon azt mondhatjuk, hogy a kísérlettel kapcsolatos egyes események meghatározott, egymáshoz viszonyított relatív súllyal következnek be, és ha valamelyik eseménnyel kapcsolatban ezt a számot önkényesen rögzítjük, akkor minden további eseményhez tartozó számérték egyértelműen meghatározottá válik. *Az* A *eseményhez ily módon tartozó számértéket az* A *esemény valószínűségének nevezzük.*

Megállapodunk abban, hogy a biztos esemény valószínűsége 1. (Azaz annak a valószínűsége, hogy a kísérlet eredménye eleme az Ω teljes eseménytérnek, éppen 1.) Ha éppen n számú kísérletet végzünk, akkor a biztos esemény gyakorisága $k_B=n$. Abból, hogy a k_A/k_B hányados stabilitást mutat következik, hogy a k_A/n hányados is stabilitást mutat.

Egy A eseménnyel kapcsolatban a k_A gyakoriságnak és a kísérletek számának (n) hányadosát, a k_A/n értéket az A esemény relatív gyakoriságának nevezzük. Gondoljuk végig, hogy ezzel a valószínűség skálája rögzített. Ugyanis a megállapodás szerint az Ω biztos esemény valószínűsége 1, az A esemény k_A/n relatív gyakorisága az A esemény valószínűségének közelítő értéke. Ha az A esemény valószínűségét $P(A)$-val jelöljük, az Ω eseményét $P(\Omega)$-val, ami a megállapodás szerint $P(\Omega)=1$, akkor minden esemény valószínűsége 0 és 1 közé esik

$$0 \leqq P(A) \leqq 1.$$

2.2. A valószínűség axiómái, valószínűségeloszlás, valószínűségi mező

Legyen A és B két egymást kizáró esemény. (Ez fontos, mert nem minden eseménypár zárja ki egymást. Gondoljunk csak arra, hogy az a két esemény, hogy a kockadobás eredménye kettőnek hatványa, ill. páros, nem zárják ki egymást.) Ha az A és B két egymást kizáró esemény, azt úgy jelöljük, hogy $AB=0$. Ekkor

$$k_{A+B} = k_A + k_B;$$

Ebből következik, hogy

$$\frac{k_{A+B}}{n} = \frac{k_A}{n} + \frac{k_B}{n}$$

és így

$$P(\mathrm{A}+\mathrm{B}) = P(\mathrm{A})+P(\mathrm{B}).$$

Egymást kizáró események valószínűsége tehát egyenlő az egyes valószínűségek összegével. Ha két esemény helyett tetszőlegesen sok, egymást kizáró eseményt tekintünk, az összegesemény valószínűsége akkor is egyenlő az egyes valószínűségek összegével.

A valószínűségszámítás axiómáit ezek után a következőképpen értelmezhetjük: legyen adott az eseménytér, értelmezzük ennek részhalmazain, az $\mathscr{A}$ halmaz A eseményein a $P(\mathrm{A})$ függvényt. Ezt valószínűségnek nevezzük, ha erre teljesülnek az alábbi axiómák:

I. $0 \leqq P(\mathrm{A}) \leqq 1,$

II. $P(\Omega) = 1.$

III. Ha $\mathrm{A}_1, \mathrm{A}_2, \ldots$

egymást páronként kizáró eseményekből álló véges vagy végtelen sorozat, vagyis ha $\mathrm{A}_i\mathrm{A}_k = \emptyset$ ha $i \neq k$, akkor

$$P\big(\sum_k \mathrm{A}_k\big) = \sum_k P(\mathrm{A}_k).$$

$P(\mathrm{A}_i)$ jelöli tehát az A_i esemény *valószínűségét*. A $P(\mathrm{A}_1)$, $P(\mathrm{A}_2), \ldots$ az $\mathrm{A}_1, \mathrm{A}_2, \ldots$ eseményrendszer *valószínűségeloszlása.*

A $P(\mathrm{A})$ függvényt ekkor az $\mathrm{A}_1, \mathrm{A}_2, \ldots$ eseményrendszer *valószínűségeloszlásának* nevezzük. (Így most a $P(\mathrm{A})$ függvény is, és $P(\mathrm{A})$-val jelöljük a konkrét számot is, ami mint valószínűség az A eseményhez tartozik. Ez nem okoz félreértést, hiszen közönséges függvényeknél is $y(x)$-szel jelöljük a függvényt is, és a konkrét számot is, amit a függvény az x független változó értékhez rendel).

Valószínűségi változó

Az olyan változót, amelynek aktuális értéke a véletlentől függ, valószínűségi változónak nevezzük. Ez így meglehetősen vulgáris, de szemléletes megfogalmazás.

Valószínűségi változók eloszlásai

Minden ξ valószínűségi változó létesít a számegyenesen egy valószínűségeloszlást, abban az értelemben, hogy a számegyenes minden részhalmazához hozzá lehet rendelni azt a valószínűséget, amellyel a valószínűségi változó számértéke az illető részhalmazba esik. Legyen X a számegyenes egy rögzített pontja, és tekintsük az

$$F(X) = P(\xi < X)$$

valószínűséget. Ha az X-et $-\infty$-től $+\infty$-ig futtatjuk, akkor egy x-től függő függvényt kapunk. Ezt a függvényt a ξ valószínűségi változó eloszlásfüggvényének nevezzük.

Arra a kérdésre tehát, hogy mi a valószínűsége annak, hogy egy valószínűségi változó egy adott értéknél kisebb (vagy nagyobb) értéket vesz fel, az eloszlásfüggvény ad választ.

Az eloszlásfüggvény egyszerű tulajdonságai a következők:

1) $F(x_1) \leqq F(x_2)$ ha $x_1 < x_2$ azaz az eloszlásfüggvény monoton.

2) $\lim\limits_{x \to -\infty} F(x) = 0$

3) $\lim\limits_{x \to +\infty} F(x) = 1.$

A valószínűségek az eloszlásfüggvényekből származtathatók.
Mivel

$$P(\xi < a) + P(a \leqq \xi < b) = P(\xi < b), \quad \text{ha} \quad a < b,$$

ezért

$$P(a \leqq \xi < b) = F(b) - F(a).$$

Eloszlások osztályozása, sűrűségfüggvény

A valószínűségi változóknak és eloszlásaiknak több fajtáját különböztethetjük meg.

a) Diszkrét valószínűségi változó. Diszkrétnek nevezzük a ξ valószínűségi változót és annak eloszlását, ha lehetséges értékei egy véges vagy végtelen $x_1, x_2, \ldots$ sorozatot alkotnak. Ekkor az $F(x)$ eloszlásfüggvény helyett szívesen dolgozunk a $p_n = P(\xi = x_n)$, $n = 1, 2, \ldots$ valószínűségekkel. Ezek egyértelműen meghatározzák a ξ eloszlását.

$$F(x) = P(\xi < x) = \sum_{x_k < x} P(\xi = x_k).$$

b) Folytonos eset. Folytonosnak nevezzük a ξ valószínűségi változót és annak eloszlását, ha van olyan $f(x) \geqq 0$ függvény, hogy a számegyenesen minden (a, b) intervalluma esetén

$$F(b) - F(a) = P(a \leqq \xi < b) = \int_a^b f(x)\, dx.$$

Az $f(x)$ függvényt a ξ valószínűségi változó *sűrűségfüggvényének* nevezzük. Az a és b határok lehetnek $-\infty$-nel és $+\infty$-nel is egyenlők, az $a = -\infty$, $b = x$ esetben az előző képlet éppen az eloszlásfüggvényt adja

$$F(x) = \int_{-\infty}^{x} f(t)\, \mathrm{d}t.$$

Az egész számegyenesen integrálva pedig

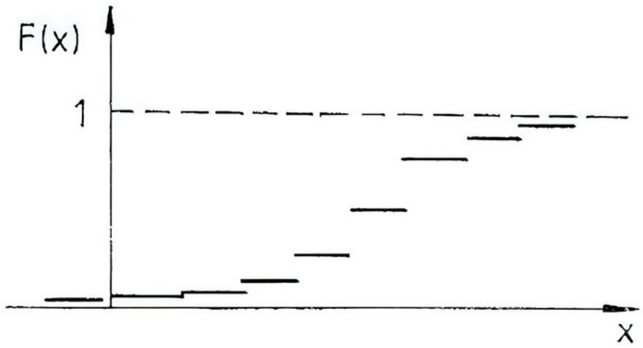
$$\int_{-\infty}^{+\infty} f(x)\, \mathrm{d}x = 1.$$

Fel kell hívni a figyelmet itt néhány fontos dologra. Folytonos eloszlás esetén az $F(x)$ az x változó folytonos függvénye, ξ tehát minden értéket 0 valószínűséggel vesz fel! Ha $F(x)$ nemcsak folytonos, de differenciálható is, akkor deriváltja éppen a sűrűségfüggvény

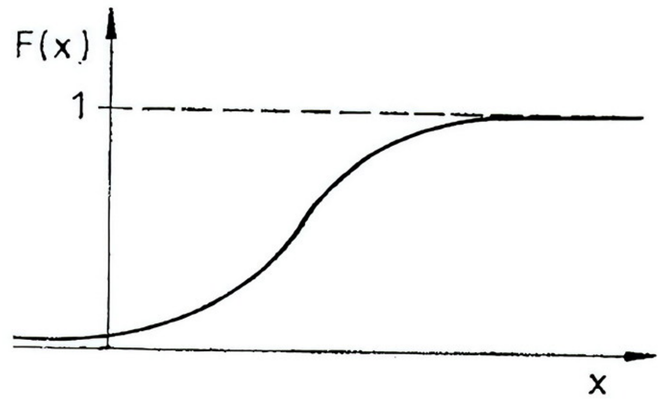
$$f(x) = F'(x).$$

Végül hangsúlyozni kívánjuk, hogy sűrűségfüggvénye csak folytonos valószínűségi változónak van.

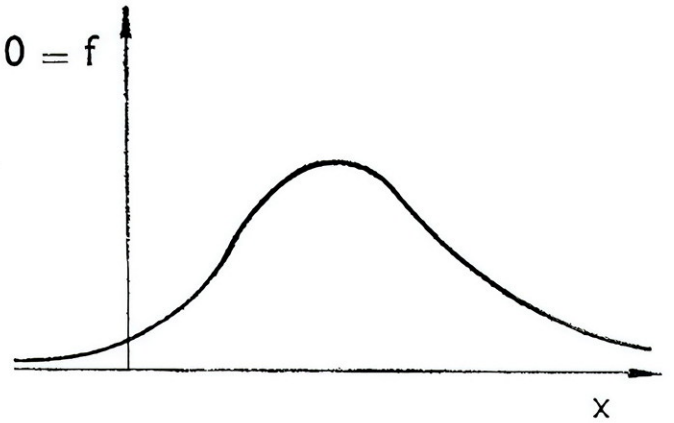
A szemléletesség kedvéért a következő ábrák a diszkrét és folytonos eset eloszlásfüggvényeit, ill. a folytonos eloszlású valószínűségi változó sűrűségfüggvényét mutatják:



1. ábra. Diszkrét valószínűségi változó eloszlásfüggvénye



2. ábra. Folytonos valószínűségi változó eloszlásfüggvénye



3. ábra. Sűrűségfüggvény

c) Kevert eset. A gyakorlatban előfordul olyan valószínűségi változó is, amelynek eloszlása az előbbi két típus keverékének tekinthető. Bizonyos $x_1, x_2, \ldots$ értékeket pozitív valószínűséggel felvesz, de vannak az eloszlásnak folytonos részei is. Pontosabban van olyan $f(x) \geqq \emptyset$ függvény, amelynek az egész számegyenesen vett integrálja véges és pozitív

$$P(a \leqq \xi < b) = \int_a^b f(x)\,\mathrm{d}x + \sum_{a \leqq x_k < b} P(\xi = x_k).$$

2.3. Várható érték, szórás

A várható érték értelmezése a valószínűség és a valószínűségi változó értelmezésén alapul. Célszerű különválasztani a diszkrét, folytonos és kevert valószínűségi változók esetét. Jelöljük $M(\xi)$-vel a ξ valószínűségi változó várható értékét. Ekkor:

a) Diszkrét eset
Ha ξ lehetséges értékei $x_1, x_2, \ldots x_n$ és ezeket $p_1, p_2, \ldots, p_n$ valószínűségekkel veszi fel, akkor

$$M(\xi) = \sum_i x_i p_i.$$

Végtelen sok x_i esetén a definíció csak akkor értelmes (azaz a várható érték csak akkor létezik), ha az előbbi képlet jobb oldalán álló sor abszolút konvergens, azaz ha

$$\sum_i |x_i| p_i < +\infty.$$

Ha ξ=állandó, akkor $p(\xi = c) = 1$, tehát $M(\xi) = c$, azaz konstans várható értéke önmaga.

b) Folytonos eset
Ha ξ folytonos eloszlású, akkor van sűrűségfüggvénye $f(x)$. Ekkor

$$M(\xi) = \int_{-\infty}^{+\infty} x f(x)\, \mathrm{d}x.$$

A diszkrét esethez hasonlóan fel kell tételezni, hogy az integrál abszolút konvergens, azaz

$$\int_{-\infty}^{+\infty} |x| f(x)\, \mathrm{d}x < +\infty,$$

ellenkező esetben azt mondjuk, hogy nincs várható értéke ξ-nek.

c) Kevert eset
Ha ξ kevert eloszlású valószínűségi változó és $p_i = P(\xi = x_i)$, akkor

$$M(\xi) = \int_{-\infty}^{+\infty} x f(x)\, \mathrm{d}x + \sum_i x_i p_i$$

feltéve, hogy

$$\int_{-\infty}^{+\infty} |x| f(x)\, \mathrm{d}x + \sum_i (x_i) p_i < +\infty,$$

különben a várható értéket nem értelmezzük.

Mivel a várható érték nem függ a valószínűségi változó konkrét jelentésétől, csupán eloszlásától, ezért a várható értéket a valószínűségi változó által a számegyenesen létesített eloszlás várható értékének nevezzük. Csupán a várható érték még eléggé kevés információt ad egy valószínűségi változó eloszlására. Adott várható értékű valószínűségi változó eloszlása még igen sokféle lehet. Pl. minden 0 pontra szimmetrikus eloszlású valószínűségi változó várható értéke 0. Sőt vannak olyan valószínűségi változók, amelyek eloszlása nem is szimmetrikus a 0 pontra várható értékük mégis 0.

A várható érték egy ξ valószínűségi változó eloszlásának centrumát adja meg. A *szórás* az eloszlás centrum körüli szóródását hivatott mérni.

A ξ valószínűségi változó szórása a $\xi - M(\xi)$ valószínűségi változó négyzetének várható értékéből vont pozitív négyzetgyök, jele $\sigma(\xi)$,

$$\sigma(\xi) = \sqrt{M\{(\xi - M(\xi))^2\}}.$$

Ahhoz, hogy szórásról beszélhessünk, eleve fel kell tételezni, hogy ξ várható értéke létezik és

$$M\{(\xi - M(\xi))^2\} + M^2(\xi) < \infty.$$

A szórás kifejezhető a ξ^2 és ξ valószínűségi változók várható értékeivel is

$$\sigma^2(\xi) = M(\xi^2) - M^2(\xi).$$

2.4. Fontosabb eloszlások

2.4.1. Diszkrét eloszlások

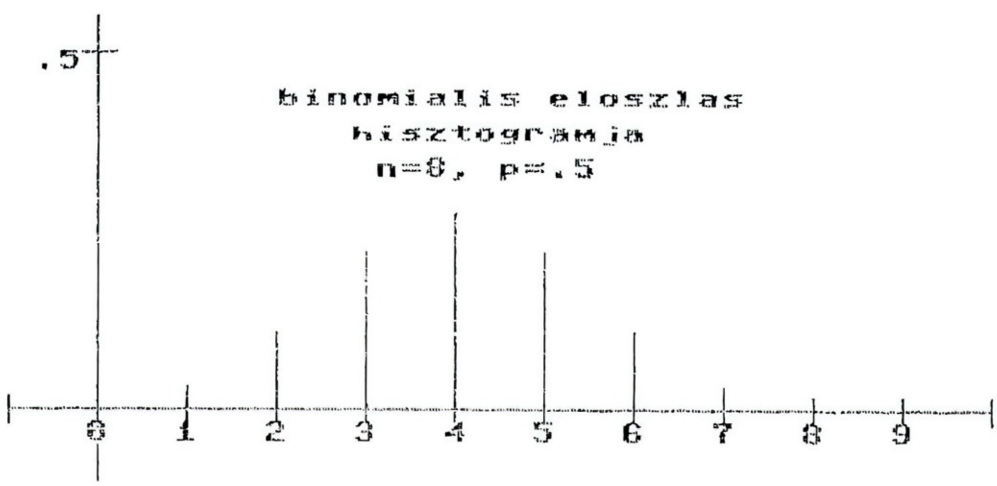
Binomiális eloszlás

Tegyük fel, hogy n független kísérletet végzünk, és a kísérletnek két lehetséges kimenetele van: a sikeres, amely p valószínűséggel következik be és a sikertelen, $1-p$ valószínűséggel. A sikeres kísérletek száma $\emptyset$ és n közé esik. Belátható, hogy annak a valószínűsége, hogy a kísérletből éppen k számú legyen sikeres,

$$P(x = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Ezzel a kifejezéssel tulajdonképpen egy diszkrét eloszlást definiáltunk, ez az ún. binomiális eloszlás.
A binomiális eloszlás két paramétere n és p.
Várható értéke $M(\xi) = np$.
Szórása $\sigma^2(\xi) = np(1-p)$.
Az eloszlás csak akkor szimmetrikus, ha $p = 0{,}5$. Egy ilyen esetet szemléltet a

FŐMENÜ-ből kiválasztott
ELOSZLÁSOK menü
ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA almenüjének
BINOMIÁLIS ELOSZLÁS programja ($n=8$, $p=0{,}5$)



4. ábra. A binomiális eloszlás hisztogramja $n=8$, $p=0{,}5$ esetén

(A program behívását, elindítását és a programcsomag használatát részletesen ismertetjük az 1. fejezetben.)

A binomiális eloszlás használatának mélyebb megértéséhez tekintsük a következő példát:

3 kockával dobunk, vagyis a kísérletek száma 3. Legyen célunk a 6-os dobás, azaz nevezzük sikeresnek azt az eseményt, amikor 6-ost dobtunk valamelyik kockával. Ennek valószínűsége nyilván bármely kocka esetén 1/6. Lehetséges, hogy egyik kockával sem dobunk hatost, ez nem sikeres ($k=\emptyset$) lehet, hogy egy kockával ($k=1$), lehet, hogy kettővel ($k=2$) és lehet, hogy mindhárommal hatost dobunk ($k=3$). Számítsuk ki, melyik esemény milyen valószínűséggel következik be a fenti négy esemény közül.

Mivel 3 kockával dobunk, $n=3$, és az előbbiek szerint $p=1/6$. Ki kell számítanunk a valószínűségeket $k=0, 1, 2, 3$ esetekre.

Válasszuk ki tehát az ELOSZLÁSOK menüjéből a BINOMIÁLIS ELOSZLÁS-t. Erről kérhetünk ismertetést, vagy egyenesen rátérhetünk a számolásra. Adjuk meg az eloszlás paramétereit és a figyelembe veendő sikeres események számát.

Első esetben	$K = 0$, $N = 3$, $p = 0{,}16666$	$P(0) = 0{,}578 \approx 0{,}58$.
Második esetben	$K = 1$, $N = 3$, $P = 0{,}16666$	$P(1) = 0{,}347 \approx 0{,}35$.
Harmadik esetben	$K = 2$, $N = 3$, $P = 0{,}16666$	$P(2) = 0{,}0694 \approx 0{,}07$.
Negyedik esetben	$K = 3$, $N = 3$, $P = 0{,}16666$	$P(3) = 0{,}0046 - 0{,}005$.

A negyedik számolás után, ha minden lehetőséget korrektül számba vettünk, az eddig kiszámolt valószínűségek összege 1,00... lesz, természetesen a kerekítési hibáktól eltekintve.

Megjegyzés: a binomiális eloszlás $p=0{,}5$ valószínűségű, két lehetséges kimenetelű, egymástól független véletlen eseményekre igen nagy N értékekre ($N>1000$) jól közelíthető a normális eloszlással (l. Folytonos eloszlások).

BINOMIÁLIS ELOSZLÁS

A változók listája:

K — a valószínűségi változó aktuális értéke,
N, P — az eloszlás paraméterei,
S — részeredmények,
J — kiszámított valószínűség,
T — kiszámított valószínűségek összege,
A$ — menünél választás,
ID, O — kiíratásnál választás (képernyő vagy printer).

Programmagyarázat:

100:	Programcsomagszervezés.
110—270:	A program neve.
280—530:	A program által nyújtott lehetőségek rövid ismertetése.
540—800:	Adatbevitel, számítások.
810—1000:	Eredmény kiírása választás szerint képernyőre vagy nyomtatóra.
1010:	Visszatérés a FŐMENÜ c. programra.
1020—1090:	Várakozás Igen-Nem válaszra.

```
540 REM****************************
550 REM*                          *
560 REM*       SZAMITASOK         *
570 REM*                          *
580 REM****************************
590 T=0
600 PRINT CHR$(147)
610 PRINT"ADJA MEG A VALTOZOK ES PARAMETEREK ERTEKEIT!"
620 PRINT
630 INPUT"K =";K
640 IF K<0 THEN PRINT"K POZITIV SZAM!":GOTO 630
650 IF K<>INT(K) THEN PRINT"K EGESZ SZAM!":GOTO 630
660 INPUT"N =";N
670 IF N<0 THEN PRINT"N POZITIV SZAM!":GOTO 660
680 IF N<>INT(N) THEN PRINT"N EGESZ SZAM!":GOTO 660
690 IF N>99999 THEN PRINT"N MAX. ERTEKE  99999!":GOTO 660
700 IF K>N THEN PRINT"K NEM LEHET NAGYOBB N-NEL!":GOTO 630
710 INPUT"P =";P
720 IF P<=0 OR P>=1 THEN PRINT"P 0 ES 1 KOZE ESIK!":GOTO 710
730 S=1
740 IF (N=0)OR(K=0)OR(K=N) THEN 780
750 FOR I=1 TO K
760 S=S*(N-I+1)/I
770 NEXT I
780 S=LOG(S)
790 J=EXP(S+K*LOG(P)+(N-K)*LOG(1-P))
800 T=T+J
```

Poisson eloszlás

A másik fontos diszkrét eloszlástípus a Poisson eloszlás. A Poisson eloszlású valószínűségi változó eloszlásfüggvénye a következő képlettel adható meg:

$$P(r) = \frac{e^{-\lambda}\mu^r}{r!} \quad (r = 0, 1, 2, \ldots, \mu > 0).$$

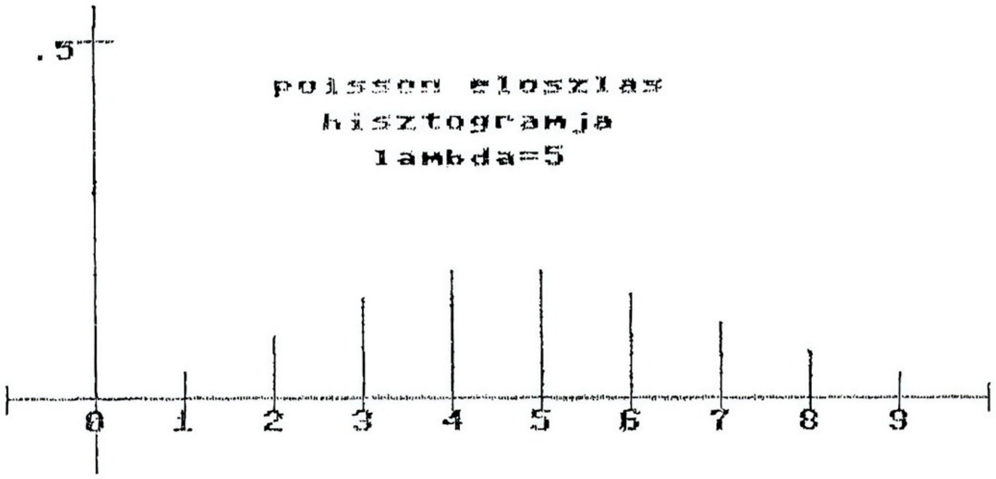
A valószínűségek természetesen pozitív értékűek, és összegük ($r=0$-tól végtelenig értve az összegzést) éppen 1. Egyetlen jellemző paramétere μ. Bizonyítható ugyanis, hogy a Poisson eloszlás várható értéke $M(X)=\lambda$, szórásnégyzete $\sigma^2(X) = \lambda$. Legegyszerűbben a következő rekurziós formulával lehet kiszámítani:

$$P(0) = e^{-\lambda},$$

$$P(r+1) = \frac{\lambda P(r)}{r+1}.$$

A Poisson eloszlásnak két fontos alkalmazási területe van: egyrészt használják az egy adott időintervallumban véletlenszerűen bekövetkező események (pl. balesetek) számának becslésére, másrészt egyszerű eszköz a binomiális eloszlás közelítésére, ha a binomiális eloszlás p paramétere (a kedvező kimenetel valószínűsége) kicsi.

A Poisson eloszlással való ismerkedés érdekében rajzoltassuk fel a képernyőre jellegzetes képét! A rajzolást a FŐMENÜ ELOSZLÁSOK menüjének ELOSZLÁS-FÜGGVÉNYEK KIRAJZOLÁSA almenüjének (8) választásával, és ott a POISSON ELOSZLÁS (2) választásával érhetjük el. A felrajzolt kép az 5 várható értékű Poisson eloszlás hisztogramjának jellegzetes képe.



5. ábra. A Poisson eloszlás hisztogramja

A Poisson eloszlás használatának illusztrálására tekintsük a következő példát: Tegyük fel, hogy egy adott ásványban a meddő részecskék eloszlása véletlenszerű, és a Poisson eloszlást követi, tapasztalatok szerint pedig átlagosan 6 meddő szemcse van egységnyi tömegű ásványban. Automatikus mintavevő berendezéssel egységnyi tömegű ásványt összegyűjtve, megszámláljuk a meddő szemcséket. Mi a valószínűsége annak, hogy 2-nél kevesebb meddő szemcsét találunk az ásványban?

$$P(X < 2) = P(X = 0) + P(X = 1).$$

A számolást természetesen kézzel is elvégezhetjük, de gyakorlásképpen hívjuk meg az ELOSZLÁSOK menü POISSON ELOSZLÁS programját (2).

A képernyőn megjelenik a használandó számítási formula és a következő felirat

ADJA MEG A VÁLTOZÓK ÉS PARAMÉTEREK ÉRTÉKEIT:
$K = ?$
$L = ?$

A kiírt képletből egyértelműen látszik, hogy K a vizsgált esetben az egységnyi tömegű ásványban feltételezett meddő szemcsék száma, azaz $K=0$-ra és $K=1$-re kell elvégezni a számítást. $K=0$ és $L=6$ paraméterezéssel

$$P(X = 0) = 2{,}47875 \cdot 10^{-3},$$

ill.

$K = 1$ és $L = 6$ paraméterezéssel

$$P(X = 1) = 0{,}014872.$$

Ezek meghatározása után, az eddig kiszámolt valószínűségek összege:

$$P(X < 2) = 0{,}01735.$$

POISSON ELOSZLÁS

A változók listája:

K — a valószínűségi változó aktuális értéke,
L — az eloszlás paramétere,
I — segédváltozó,
J — kiszámított valószínűség,
T — kiszámított valószínűségek összege,
A$ — menünél választás,
ID, O — kiíratásnál választás (képernyő vagy printer).

Programmagyarázat:

100:	Programcsomag-szervezés.
110—270:	A program neve.
280—510:	A program által nyújtott lehetőségek rövid ismertetése.
520—740:	Adatbevitel, számítások.
750—940:	Eredmény kiírása választás szerint képernyőre vagy nyomtatóra.
950:	Visszatérés a „FŐMENÜ” c. programra.
960—1030:	Várakozás Igen-Nem válaszra.

```
520 REM*****************************
530 REM*                           *
540 REM*       SZAMITASOK          *
550 REM*                           *
560 REM*****************************
570 T=0
580 PRINT CHR$(147)
590 PRINT"ADJA MEG A VALTOZOK ES PARAMETEREK ERTEKEIT!"
600 PRINT
610 INPUT"K =";K
620 IF K<0 OR K<>INT(K) THEN PRINT"K POZITIV EGESZ SZAM!":GOTO 610
630 IF K>999 THEN PRINT"K MAXIMALIS ERTEKE 999":GOTO 610
640 PRINT
650 INPUT"L =";L
660 IF L<0 THEN PRINT"LAMBDA POZITIV SZAM!":GOTO 650
670 J=0
680 IF K=0 THEN 730
690 FOR I=1 TO K
700 J=J+LOG(I)
710 NEXT I
730 J=EXP(-L+K*LOG(L)-J)
740 T=T+J
```

2.4.2. Folytonos eloszlások

Normális eloszlás (Gauss eloszlás)

A valóság véletlen jelenségeinek modellezése során gyakran élünk azzal a feltevéssel, hogy a valószínűségi változó normális eloszlású. A feltevés azon a matematikai tételen alapszik, miszerint nagyszámú, egymástól független valószínűségi változó összege az egyes változók eloszlásától függetlenül mindig normális eloszlású, ha az egyes változók véletlen ingadozása az összeg ingadozásához képest kicsi. Tapasztalat szerint az arányos skálán mért mennyiségek véletlen mérési hibái igen gyakran a normális eloszlás törvényét követik.

Folytonos eloszlás, szórása és várható értéke egyértelműen meghatározza. A könnyebb kezelhetőség kedvéért standardizálják $M=\emptyset$ és $\sigma=1$ paraméterűvé. Eloszlásfüggvénye (és sűrűségfüggvénye) csak az M-től és σ-tól függ. Ezért ha a standardizált alakot használjuk, ez már csak a valószínűségi változó értékétől (x) függ. Ez teszi éppen lehetővé a táblázatos megadást és az egységes átszámolást.

Jellegzetes képét ki is rajzoltathatjuk a képernyőn, ill. a sornyomtatón. Ehhez (feltételezve, hogy a statisztikai programcsomag FŐMENÜ nevű programját már betöltöttük és a RUN paranccsal elindítottuk) a következőket kell tennünk:

— Válasszuk ki a menüből az ELOSZLÁSOK programágat a (3) gomb benyomásával. Ekkor megjelenik a képernyőn a következő felirat:

TEGYE BE A LEMEZ "A" OLDALÁT!

— Ha a lemez a meghajtóban volt, a RETURN gomb megnyomására megjelenik az ELOSZLÁSOK program almenüje. Ebből a (8) gomb lenyomásával kiválasztjuk az ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA (8) programágat.
— A képernyőn megjelenik az ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA almenüje. Ebből a NORMÁLIS ELOSZLÁS kirajzolását a (3) gomb megnyomásával választhatjuk ki.

A program először a képernyőre rajzolja fel a standard normális eloszlás eloszlás- és sűrűségfüggvényét (6. ábra). Jól megfigyelhetjük a sűrűségfüggvény jellegzetes haranggörbe alakját, és az eloszlásfüggvény tipikus S görbéjét. Az előbbi az Y tengelyre szimmetrikus, az utóbbi a (0; 0,5) pontra. A görbék felrajzolása után a képernyőn a következő kérdés jelenik meg.

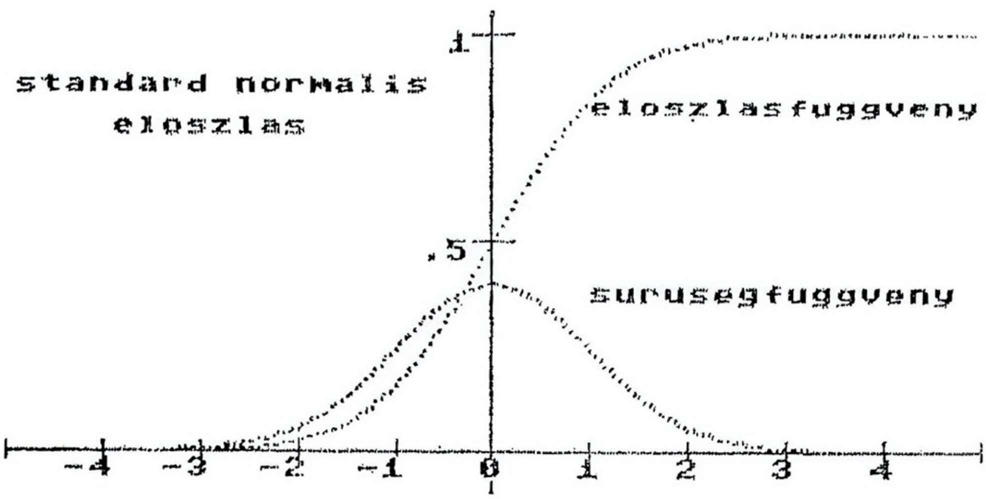
KIRAJZOLJUK PRINTERRE IS? (I/N)

Az I betű gombjának lenyomására a sornyomtatóra is kirajzolódik mindkét függvény.

A normális eloszlásról többet is megtudhatunk a programcsomag segítségével. Ha ugyanis az ELOSZLÁSOK menüjéből nem az eloszlásfüggvények kirajzolása ágat választjuk, hanem a

NORMÁLIS ELOSZLÁS (3)

menüágat, akkor lehetőség van a függvények jellemző értékeinek kiírására is. Pontosabban mind standardizált, mind nem standardizált normális eloszláshoz megadhatjuk a valószínűségi változó értékét, és ehhez a program megadja a sűrűség- és eloszlásfüggvény értékeket. Ennek jelentőségét megérthetjük, ha arra gondolunk, hogy a valószínűségek az eloszlásfüggvény-értékekből származtathatók. Olyan kérdésekre lehet így választ kapni, mint pl.:



6. ábra. A normális eloszlás sűrűség- és eloszlásfüggvénye

— mi a valószínűsége annak, hogy egy adott várható értékű és szórású normális eloszlású valószínűségi változó értéke kisebb, mint a várható érték és a szórás összege, a várható érték és a szórás kétszeresének összege, a várható érték és a szórás háromszorosának összege.

Válaszoljuk meg ezeket a kérdéseket! Legyen a várható érték $M(\xi)=10$ és a szórás $D(\xi)=2$. Ha nem kértünk a programtól ismertetést, akkor megjelenik a képernyőn a következő kérdés:

STANDARDIZÁLT (S) VAGY NEM STANDARDIZÁLT (N) VÁLTOZÓRÓL VAN SZÓ?

Az (N) billentyű lenyomására a program kéri a változó várható értékét $(M(\xi))$ és szórását $(D(\xi))$, majd a valószínűségi változó értékét (X). Ez utóbbi az első esetben $10+2=12$. Ezt beütve, és a képernyőt választva megjelenítő perifériaként, kiíródik a képernyőre a válasz:

a sűrűségfüggvény értéke	0,24197...
az eloszlásfüggvény értéke	0,84135...

Ez az eredmény azt jelenti, hogy 84,13% a valószínűsége annak, hogy a valószínűségi változó értéke kisebb, mint $M(\xi)+D(\xi)$.

A három feltett kérdésre adott választ összefoglalva:

$$p(x \leqq M(\xi)+D(\xi)) = 0{,}8413,$$

$$p(x \leqq M(\xi)+2D(\xi)) = 0{,}9772,$$

$$p(x \leqq M(\xi)+3D(\xi)) = 0{,}9986.$$

Megjegyzendő, hogy a válasz nem függ $M(\xi)$ és $D(\xi)$ konkrét értékétől, azaz ugyanezt a választ kapnánk például a standardizált normális eloszlású valószínűségi változó esetére is. Másrészt gondoljuk meg, hogy az eloszlás sűrűségfüggvénye szimmetrikus és az eloszlásfüggvény a sűrűségfüggvény integrálja. Ez azt is jelenti, hogy normális eloszlású valószínűségi változó esetén a várható érték $\pm 1D(\xi)$ környezetébe esik a valószínűségi változó $(0{,}84135-0{,}5)2=68{,}27$%-os valószínűséggel, az $M(\xi)\pm 2D(\xi)$ környezetbe esik $(0{,}9772-0{,}5)2=95{,}44$%-os, és $M(\xi)\pm$ $\pm 3D(\xi)$ környezetébe esik $(0{,}9986-0{,}5)2=99{,}72$%-os valószínűséggel.

A normális eloszlásból származtatható eloszlástípusokhoz tartozik a χ^2 eloszlás, a t eloszlás és az F eloszlás is.

NORMÁLIS ELOSZLÁS

A változók listája:

M	— várható érték,
D	— szórás,
X	— a valószínűségi változó aktuális értéke,
Z	— segédváltozó,
E	— sűrűségfüggvény értéke,
Q	— eloszlásfüggvény értéke,
A$, X$, H$	— menünél választás,
ID, O	— kiíratásnál választás (képernyő vagy nyomtató).

Programmagyarázat:

100:	Programcsomag-szervezés.
110—270:	A program neve.
280—510:	A program által nyújtott lehetőségek rövid ismertetése.
520—800:	Adatbevitel, számítások.
810—1020:	Eredmény kiírása választás szerint képernyőre vagy nyomtatóra.
1030:	Visszatérés a FŐMENÜ c. programra.
1040—1110:	Várakozás Igen-Nem válaszra.

```
520 REM*****************************
530 REM*                           *
540 REM*        SZAMITASOK         *
550 REM*                           *
560 REM*****************************
570 PRINT CHR$(147)
580 PRINT"STANDARDIZALT (S)  VAGY"
590 PRINT"NEM STANDARDIZALT (N)"
600 INPUT"VALTOZOROL  VAN SZO";H$
610 PRINT
620 IF H$<>"S" AND H$<>"N" THEN 580
630 IF H$="S" THEN 700
640 PRINT:PRINT:PRINT
650 PRINT"ADJA MEG A VARHATO ERTEKET!"
660 INPUT"M = ";M
670 PRINT:PRINT"ADJA MEG A SZORAST!"
680 INPUT"D = ";D
690 GOTO 710
700 D=1:M=0
710 PRINT:PRINT
720 PRINT:PRINT"ADJA MEG A VALTOZO ERTEKET!"
730 INPUT"X = ";X
740 X=(X-M)/D
750 E=EXP(-X↑2/2)/2.5066283
760 Z=X
770 X=1/(1+.33267*ABS(X))
780 Q=1-E*(.4361836*X-.1201676*X↑2+.937298*X↑3)
790 IF Z>=0 THEN 810
800 Q=1-Q
```

Exponenciális eloszlás

Fontos folytonos eloszlás definíciója: a ξ folytonos eloszlású valószínűségi változó exponenciális eloszlású, ha sűrűségfüggvénye a következőképpen írható fel:

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{ha} \quad x>0 \quad (\lambda>\vartheta)\\ 0 & \text{ha} \quad x<0.\end{cases}$$

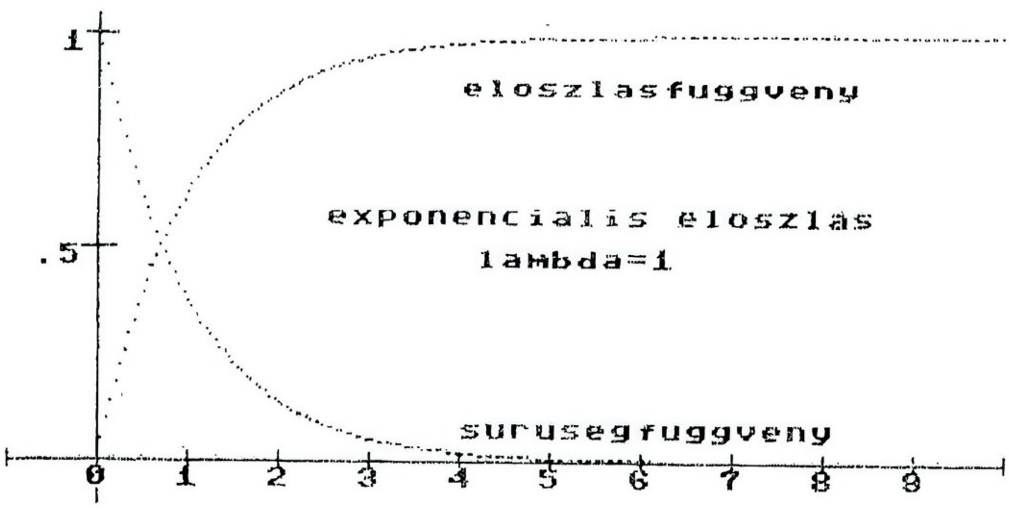
Az előbbiekből következően eloszlásfüggvénye:

$$F(x)=\begin{cases}1-e^{-\lambda x} & \text{ha} \quad x\geqq 0\\ 0 & \text{ha} \quad x<0.\end{cases}$$

Egyetlen paramétere λ. Várható értéke $1/\lambda$, szórása ugyancsak $1/\lambda$.

Az exponenciális eloszlást elsősorban bizonyos gépi berendezések várható meghibásodási idejének becslésére használják. Olyan berendezéseknél, amelyeknél a meghibásodás valószínűsége nem függ a berendezés addigi „előéletétől”. Másik lényeges alkalmazási területe a sorbanállási problémák vizsgálata.

Jellegzetes képét az ELOSZLÁSOK menü ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA (8) menüágából az EXPONENCIÁLIS ELOSZLÁS kiválasztásával a képernyőre (ill. nyomtatóra is) kirajzoltathatjuk. A 7. ábrán felrajzolt eset a $\lambda=1$ paraméterű exponenciális eloszlás sürűség- és eloszlásfüggvényét mutatja.



7. ábra. $\lambda=1$ paraméterű exponenciális eloszlás sűrűség- és eloszlásfüggvénye

Alkalmazására illusztrációképpen oldjuk meg a következő feladatot:
Egy bizonyos elektronikus alkatrész élettartamáról tudjuk, hogy exponenciális eloszlású, élettartamának várható értéke 100 óra. Az ilyen alkatrészek hány százaléka fog meghibásodni az első 10 órában, a második 10 órában, ill. a harmadik 10 órás periódusban?
Mivel az átlagos várható élettartam 100 h, $\lambda=1/100$.
A példa megoldásához az ELOSZLÁSOK menüjéből válasszuk ki az EXPONENCIÁLIS ELOSZLÁS-t. Kiválasztása után meg kell adni λ értékét $(=0{,}01)$, majd az X értéket, amelynél az eloszlásfüggvény értékét keressük. Ez az első kérdés megválaszolásakor 10. A képernyőn megjelenik az eredmény:

$$P(S < 10) = 0{,}09516,$$

azaz az első 10 órában átlagban az alkatrészek 9,52%-a hibásodik meg. Mit várhatunk a következő 10 órában?

Mivel ismét az exponenciális eloszlást fogjuk használni, arra a kérdésre, hogy AKAR ÚJABB SZÁMÍTÁST VÉGEZNI? (I/N) az I billentyű megnyomásával, igennel válaszolunk. Ekkor újra meg kell adni λ és X értékét (0,01 és 20). A képernyőn megjelenő eredmény:

A VALÓSZÍNŰSÉG: 0,1812... .

Ebből levonva az első 10 órában meghibásodott alkatrészek százalékos értékét, a második 10 órában meghibásodó alkatrészek százalékban kifejezett mennyiségét kapjuk:

$$P(10 < x \leqq 20) = 0{,}18126-0{,}09516 = 0{,}0861.$$

Azaz az alkatrészek 8,61%-a fog várhatóan meghibásodni a második 10 órában. Hasonló módon az $x=30$-hoz tartozó eloszlásfüggvény értékből (0,25918) a harmadik 10 órában meghibásodó alkatrészek aránya

$$P(20 < x \leqq 30) = 0{,}25918-0{,}18126 = 0{,}07792,$$

azaz ~7,8%.

EXPONENCIÁLIS ELOSZLÁS

A változók listája:

L — lambda paraméter értéke,
X — a valószínűségi változó értéke,
P — kiszámított valószínűség,
A\$, X\$ — menünél választás,
ID, O — kiíratásnál választás (képernyő vagy printer).

Programmagyarázat:

100:	Programcsomag-szervezés.
110—270:	A program neve.
280—630:	A program által nyújtott lehetőségek rövid ismertetése.
640—760:	Adatbevitel, számítások.
770—980:	Eredmény kiírása választás szerint képernyőre vagy nyomtatóra.
990:	Visszatérés a FŐMENÜ c. programra.
1000—1070:	Várakozás Igen-Nem válaszra.

χ^2 eloszlás

Csak pozitív valós számokra van értelmezve. Definíciója: legyenek $x_1, x_2, \ldots, x_n$ egymástól független standard (0 várható értékű és 1 szórású) normális eloszlású valószínűségi változók. Ekkor a

$$\chi^2 = x_1^2+x_2^2+\ldots+x_n^2.$$

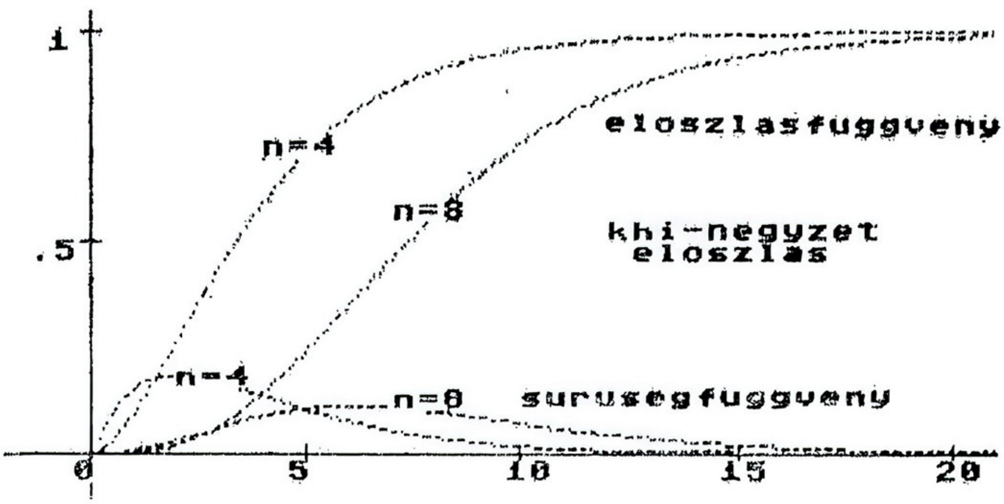
χ^2 eloszlású valószínűségi változó, n szabadsági fokkal (szabadságfok: a független összeadandók száma). Paramétere n. Várható értéke $M(\chi^2)=n$, szórásnégyzete $\sigma^2=2n$.

```
640 REM******************************
650 REM*                            *
660 REM*       SZAMITASOK           *
670 REM*                            *
680 REM******************************
690 PRINT CHR$(147)
700 PRINT"ADJA MEG A LAMBDA PARAMETER ERTEKET!"
710 INPUT"L(LAMBDA) = ";L
720 IF L<0 THEN PRINT"LAMBDA>=0 KELL LEGYEN!":GOTO 710
730 PRINT:PRINT"ADJA MEG A VALTOZO ERTEKET!"
740 INPUT"X = ";X
750 IF X<0 THEN PRINT"X>=0 KELL LEGYEN!":GOTO 740
760 F=1-EXP(-L*X)
```

A χ^2 eloszlás jobb megismeréséhez válasszuk ismét a FŐMENÜ-ből az ELOSZLÁSOK programágat (3), majd az ELOSZLÁSOK menüjéből az ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA programrészt, végül ennek almenüjéből a

KHI-NÉGYZET ELOSZLÁS-t (5).

Az (5) billentyű lenyomására kirajzolódik a képernyőre az $n=4$ és $n=8$ szabadsági fokú χ^2 eloszlás sűrűség- és eloszlásfüggvénye.



8. ábra. A χ^2 eloszlás sűrűség- és eloszlásfüggvénye $n=0, 1, 4, 8$ esetre

Látható, hogy ez az eloszlás csak pozitív x értékekre van értelmezve, és nem szimmetrikus, sőt a várható értéknél kisebb értékek valószínűsége sem 0,5 (mint azt első pillanatban várnánk). Erről könnyen meg is bizonyosodhatunk. Válasszuk ugyanis az ELOSZLÁSOK menüjéből a KHI-NÉGYZET programágat, és az ismertetés kiíratását mellőzve, válasszuk mondjuk az $n=5$ szabadsági fokú χ^2 eloszlást, és legyen a khi-négyzet eloszlású valószínűségi változó értéke éppen a várható érték, azaz 5.

A képernyőre (vagy printerre) kiíratott válasz a következő lesz:

N = 5 KHI-NÉGYZET = 5
P = Ø,5841

Ez azt jelenti, hogy a várható értéknél kisebb értékek valószínűsége $n=5$ szabadsági fokú khi-négyzet eloszlás esetén 58,41%. Ez a valószínűség természetesen függ az eloszlás szabadsági fokától. A szabadsági fok növekedtével tart a 0,5 érték felé:

Szabadsági fok	A várható értéknél kisebb értékek valószínűsége
5	0,5841
10	0,5595
20	0,542
30	0,5343
45	0,528

KHI-NÉGYZET ELOSZLÁS

A változók listája:

V — szabadsági fok,
C — khi-négyzet értéke,
J, A, B, S, T — részeredmények,
W, I — segédváltozók,
A$, X$ — menünél választás,
ID, O — kiíratásnál választás (képernyő vagy nyomtató).

Programmagyarázat:

100: Programcsomag-szervezés.
110—270: A program neve.
280—540: A program által nyújtott lehetőségek rövid ismertetése.
550—810: Adatbevitel, számítások.
810—1020: Eredmény kiírása választás szerint képernyőre vagy nyomtatóra.
1030: Visszatérés a FŐMENÜ c. programra.
1040—1110: Várakozás Igen-Nem válaszra.

```
550 REM******************************
560 REM*                            *
570 REM*        SZAMITASOK          *
580 REM*                            *
590 REM******************************
600 PRINT CHR$(147)
610 INPUT"ADJA MEG A SZABADSAGI FOKOT";V
620 IF V<=0 THEN PRINT"A SZABADSAGI FOK POZITIV!":GOTO 610
630 IF V<>INT(V) THEN PRINT"A SZABADSAGI FOK EGESZ SZAM!":GOTO 610
640 PRINT:INPUT"ADJA MEG KHI-NEGYZET ERTEKET";C
650 IF C<0 THEN PRINT"KHI-NEGYZET NEM NEMATIV!":GOTO 640
660 J=1
670 FOR I=V TO 2 STEP -2
680 J=J*I
690 NEXT I
700 A=C↑(INT((V+1)/2))*EXP(-C/2)/J
710 IF INT(V/2)=V/2 THEN 740
720 B=SQR(2/C/3.1415927)
730 GOTO 750
740 B=1
750 S=1:T=1
760 W=V
770 V=V+2
780 T=T*C/V
790 IF T<.00001 THEN 820
800 S=S+T
810 GOTO 770
```

Student eloszlás (t eloszlás)

A Student eloszlás meghatározása W. Gosset ír kémikustól származik. A következőképpen értelmezzük: legyenek $Y, x_1, x_2, \ldots, x_n$ független, standard normális eloszlású valószínűségi változók. Képezzük az n szabadsági fokú

$$\chi^2 = x_1^2 + x_2^2 + \ldots + x_n^2.$$

χ^2 eloszlású valószínűségi változót. Osszuk el a szabadsági fokok számával és vonjunk a hányadosból négyzetgyököt.

Így

$$\sqrt{\frac{\chi^2}{n}} \text{-hez jutunk.}$$

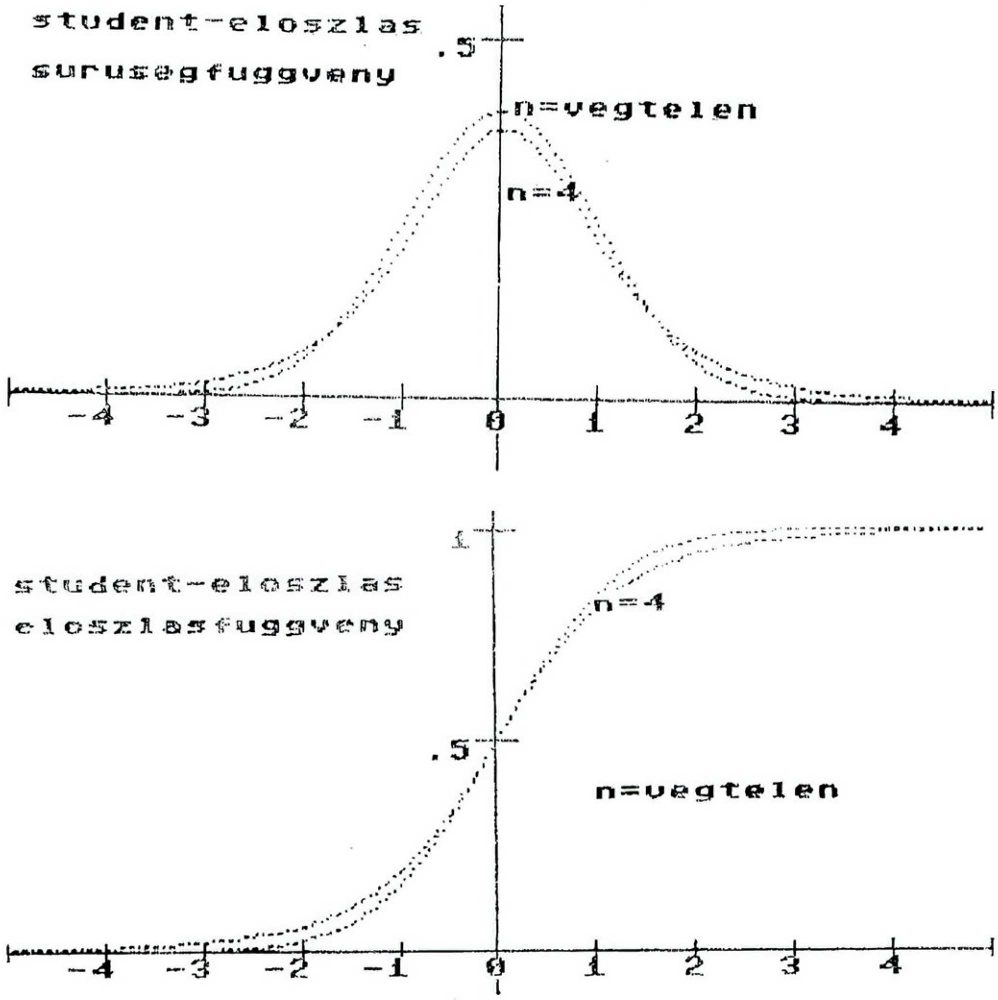
Ha az Y standard normális eloszlású valószínűségi változót osztjuk a fenti $\sqrt{\frac{\chi^2}{n}}$-nel, az n szabadsági fokú Student eloszlású valószínűségi változóhoz jutunk

$$t_n = \frac{\sqrt{n}Y}{\sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}}.$$

Értelmezési tartománya: $-\infty < x < +\infty$. Paramétere: n, ahol n pozitív egész szám. Elég jól megközelíti a normális eloszlást, különösen, ha n elég nagy. $n \geqq 20$ esetén gyakorlatilag egybeesik a normális eloszlással.

AZ ELOSZLÁSOK menü ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA almenüjének STUDENT ELOSZLÁS programjával először az $n=4$ és az $n=\infty$ szabadsági fokú Student eloszlás sűrűségfüggvényeit rajzoltathatjuk ki. Ezt mutatja a 9. ábra is.

Látható, hogy az $n=\infty$ szabadsági fokú Student eloszlás éppen a normális eloszlás. Az eloszlás sűrűségfüggvényének kirajzolása után a megfelelő eloszlásfüggvények jelennek meg a képernyőn. Mindkettő kiíratható nyomtatón is. A legegysze-



9. ábra. A Student eloszlás sűrűség- és eloszlásfüggvénye $n=4$ és $n=\infty$ szabadsági fokok esetén

rűbb kérdés, amit a Student eloszlással megfogalmazhatunk, a következő: mi a valószínűsége annak, hogy egy 0 várható értékű n szabadsági fokú Student eloszlású valószínűségi változó értéke egy előre megadott értéknél nagyobb?

Legyen feladatunk például annak kiszámítása, hogy milyen valószínűséggel lesz egy ξ 0 várható értékű, 10 szabadsági fokú Student eloszlású valószínűségi változó értéke nagyobb mint 1.

Erre a kérdésre választ kaphatunk az ELOSZLÁSOK menü STUDENT ELOSZLÁS programjának használatával (6-os menüág).

Az interaktív program először megkérdezi, kérünk-e ismertetést a programról. Ha igennel válaszolunk, megjelenik a képernyőn egy egyoldalas ismertetés. Ebből (többek között) azt is megtudhatjuk, hogy a program ún. kétoldali eloszlást számol. Ez azt jelenti, hogy a $t=1$ értékhez azt a valószínűséget számolja ki, amellyel a valószínűségi változó értéke a $[-1, 1]$ intervallumon kívül esik. Ezért a feltett kérdésre majd úgy kell helyesen válaszolnunk, hogy a kért valószínűséget 2-vel osztjuk.

Az ismertető elolvasásának nyugtázása után a program megkérdezi a szabadsági fokot. Esetünkben $n=10$. Utána megjelenik a kérés: ADJA MEG A T ÉRTÉKET!

Erre válaszképpen beírjuk az 1-et, majd a megjelenítésre kiválaszthatjuk a képernyőt (3) vagy a nyomtatót (4). Az eredmény:

$$P = 0,3427.$$

Ez úgy értendő, hogy az 1 értéknél nagyobb értékek valószínűsége

$$P(\xi \geqq 1) = 0{,}3427/2 = 0{,}1713; \quad \text{azaz kb. } 17\%.$$

STUDENT ELOSZLÁS

A változók listája:

T	— t-érték,
D	— szabadsági fok,
A, B, T1	— segédváltozók,
A1, B1, Z	— részeredmények,
A$, X$	— menünél választás,
ID, O	— kiíratásnál választás (képernyő vagy nyomtató).

Programmagyarázat:

100:	Programcsomag-szervezés.
110—270:	A program neve.
280—590:	A program által nyújtott lehetőségek rövid ismertetése.
600—920:	Adatbevitel, számítások.
930—1130:	Eredmény kiírása választás szerint képernyőre vagy nyomtatóra.
1140:	Visszatérés a FŐMENÜ c. programra.
1150—1220:	Várakozás Igen-Nem válaszra.

```
600 REM****************************
610 REM*                          *
620 REM*       SZAMITASOK         *
630 REM*                          *
640 REM****************************
650 PRINT CHR$(147)
660 PRINT"ADJA MEG A SZABADSAGI FOKOT!"
670 INPUT D
680 IF D<=0 THEN PRINT"A SZABADSAGI FOK POZITIV!":GOTO 670
690 IF D<INT(D) THEN PRINT"A SZABADSAGI FOK EGESZ":GOTO 670
700 PRINT:PRINT"ADJA MEG A T-ERTEKET!"
710 INPUT T
720 P=1:M=1
730 T=T*T
740 IF T<1 THEN 790
750 A=M
760 B=D
770 T1=T
780 GOTO 820
790 A=D
800 B=M
810 T1=1/T
820 A1=2/9/A
830 B1=2/9/B
840 Z=ABS((1-B1)*T1↑(1/3)-1+A1)/SQR(B1*T1↑(2/3)+A1)
850 IF B<4 THEN 880
860 P=.5/(1+Z*(.196854+Z*(.115194+Z*(.000344+Z*.019527))))↑4
870 GOTO 900
880 Z=Z*(1+.08*Z↑4/B↑3)
890 GOTO 850
900 IF T<1 THEN 920
910 GOTO 930
920 P=1-P
```

Fisher eloszlás (F eloszlás)

Legyen két független, χ^2 eloszlású, n_1 és n_2 szabadsági fokú valószínűségi változónk, X_1 és X_2. A kettő hányadosából képzett

$$F = \frac{X_1/n_1}{X_2/n_2}$$

valószínűségi változó eloszlása a Fisher eloszlás. Két paramétere n_1 és n_2. Várható értéke:

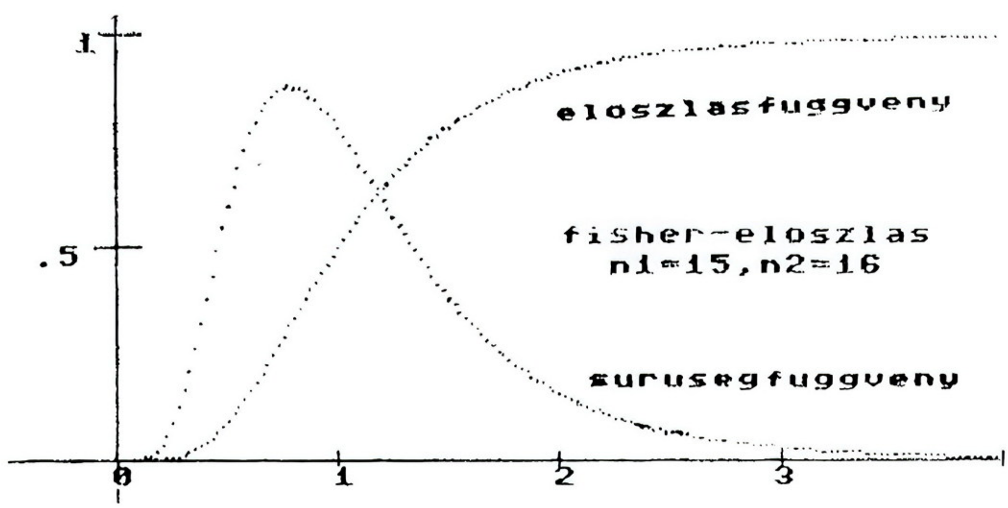
$$M(F) = \frac{n_2}{n_2-2}, \quad \text{feltéve, hogy} \quad n_2 > 2.$$

Szórásnégyzete:

$$\sigma^2(F) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}.$$

Elsősorban szórások összehasonlításánál fogjuk alkalmazni.

AZ ELOSZLÁSOK menü ELOSZLÁSFÜGGVÉNYEK KIRAJZOLÁSA almenüjének FISHER ELOSZLÁS programjával a Fisher eloszlás sűrűség- és eloszlásfüggvényének jellegzetes képét rajzoltathatjuk ki a képernyőn, ill. nyomtathatjuk. A kirajzolt függvény olyan Fisher eloszlást mutat, amelynél a számláló szabadsági foka $n_1=15$, a nevező szabadsági foka $n_2=16$. Ezt az esetet mutatjuk be a 10. ábrán is. (Az ábrán jól látható, hogy ez az eloszlás nem szimmetrikus!)



10. ábra. Fisher eloszlás sűrűség- és eloszlásfüggvénye. $n_1=15$; $n_2=16$

AZ ELOSZLÁSOK KIRAJZOLÁSA menüágból visszatérhetünk a FŐMENÜ-be.

A legegyszerűbb kérdés a Fisher eloszlás használatával kapcsolatban úgy fogalmazható meg: mi a valószínűsége annak, hogy egy adott szabadsági fokpárhoz tartozó Fisher eloszlású valószínűségi változó értéke egy adott értéket meghalad. Legyen például $n_1=10$; $n_2=8$. Ekkor az elméleti bevezetőből tudjuk, hogy a várható érték

$$M = \frac{8}{8-2} = \frac{8}{6} = 1{,}333.$$

Számítsuk ki annak a valószínűségét, hogy a mondott tulajdonságú valószínűségi változó értéke nagyobb, mint a várható érték kétszerese, azaz 2,67.

Az ELOSZLÁSOK menü FISHER ELOSZLÁS programjának kiválasztásakor először ismertetést kérhetünk az eloszlásról, majd (akár kértünk ismertetőt, akár nem) meg kell adni a képernyőre kiírt kérdésekre adott válaszok formájában a kiválasztott F értéket (2,67), a számláló szabadsági fokát (10) és a nevező szabadsági fokát (8). Az eredmény (képernyőn vagy nyomtatón megjelenítve) a következő:

N1 = 10, N2 = 8, F = 2,67

P = 0,0888.

Ezt úgy kell értelmezni, hogy 2,67-nél nagyobb értékek valószínűsége 8,88%.

FISHER ELOSZLÁS

A változók listája:

F	— f-érték,
N	— számláló szabadsági foka,
D	— nevező szabadsági foka,
R, B, F1	— segédváltozók,
R1, B1, Z	— részeredmények,
P	— kiszámított valószínűség,
A$	— választás ismertetés kérésnél, ill. további számolásnál.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—281:	A program neve.
290—510:	A program által nyújtott lehetőségek rövid ismertetése.
520—850:	Az adatok bevitele, számítások.
860—1060:	Az eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
1070:	Visszatérés a FŐMENÜ c. programra.
1080—1150:	Várakozás Igen-Nem válaszra.

```
520 REM****************************
530 REM*                          *
540 REM*       SZAMITASOK         *
550 REM*                          *
560 REM****************************
570 PRINT CHR$(147)
580 INPUT"AZ F-ERTEK";F
590 IF F<=0 THEN PRINT"AZ F-ERTEK POZITIV!":GOTO 580
600 INPUT"A SZAMLALO SZABADSAGI FOKA";N
610 IF N<=0 THEN PRINT"A SZABADSAGI FOK POZITIV!":GOTO 600
620 IF N<>INT(N) THEN PRINT"A SZABADSAGI FOK EGESZ SZAM!":GOTO 600
630 INPUT"A NEVEZO SZABADSAGI FOKA";D
640 IF D<=0 THEN PRINT"A SZABADSAGI FOK POZITIV!":GOTO 630
650 IF D<>INT(D) THEN PRINT"A SZABADSAGI FOK EGESZ SZAM!":GOTO 630
660 P=1
670 IF F<1 THEN 720
680 A=N
690 B=D
700 F1=F
710 GOTO 750
720 A=D
730 B=N
740 F1=1/F
750 A1=2/9/A
760 B1=2/9/B
770 Z=ABS((1-B1)*F1↑(1/3)-1+A1)/SQR(B1*F1↑(2/3)+A1)
780 IF B<4 THEN 810
790 P=.5/(1+Z*(.196854+Z*(.115194+Z*(.000344+Z*.019527))))↑4
800 GOTO 830
810 Z=Z*(1+.08*Z↑4/B↑3)
820 GOTO 790
830 IF F<1 THEN 850
840 GOTO 860
850 P=1-P
```

3. MATEMATIKAI STATISZTIKAI ALAPFOGALMAK

3.1. Statisztikai sokaság, minta

A statisztika tömegesen előforduló jelenségek jellemzése minőségi vagy mennyiségi adatokkal. Minőségi jellemzéssel ezen a helyen nem foglalkozunk. Az egyszerűség kedvéért tételezzük fel, hogy véletlen kísérletet végzünk, és megfigyelt kimenetelét egyetlen számadattal jellemezzük. A kísérletek elvégzésével és kimenetelük megfigyelésével tulajdonképpen *mintát* veszünk az összes lehetséges kimenetelek halmazából. Ez utóbbit gyakran *statisztikai sokaságnak* nevezik. Mintát tulajdonképpen azért veszünk, mert igen költséges lenne minden lehetséges kísérletet elvégezni. A statisztikakészítés alapgondolata lényegében az, hogy a minta valamiképpen jellemző a sokaságra, és lehetőséget ad a sokaságra vonatkozó becslések elvégzésére. A mintavétel a sokaság n számú elemének véletlenszerű kiválasztásából áll. Jelöljék $x_1, x_2, \ldots, x_n$ a kiválasztott elemekhez tartozó számértékeket a kiválasztás sorrendjében. A kiválasztás minden egyes lépésben véletlenszerű, ezért bár mindig n konkrét számot kapunk, kaphattunk volna másokat is, ezért az $x_1, x_2, \ldots, x_n$ valószínűségi változók.

Statisztikai vizsgálatot a sokaság eloszlására vonatkozó teljes vagy részleges információ szerzése érdekében végzünk. Lehetséges, hogy a sokaság eloszlását, de lehetséges, hogy ennek csak egyes paramétereit akarjuk meghatározni. Ha ismerjük a statisztikai sokaság eloszlásának típusát, de nem ismerjük az eloszlás analitikus kifejezésében szereplő paraméterek konkrét értékeit, akkor ezeket ún. statisztikai becsléssel lehet az n elemű minta alapján közelíteni.

Ha feltételezzük, hogy csak egy paramétert kell becsülnünk, a sokaság eloszlásfüggvénye tehát egy $F(x, a)$ függvény, akkor az a valós állandó értékét becsülhetjük az $x_1, x_2, \ldots, x_n$ mintaelemek egy

$$\hat{a} = \hat{a}(x_1, x_2, \ldots, x_n) \text{ függvénye alapján.}$$

Az $x_1, x_2, \ldots, x_n$ mintabeli változók egy tetszőleges függvényét statisztikai függvénynek vagy röviden statisztikának nevezzük. Az a paraméter közelítésére alkalmazott statisztikát az a paraméter becslésének nevezzük, és az általa nyert konkrét â értéket az a paraméter becsült értékének nevezzük.

3.2. A minta feldolgozása: rendezett minta, osztályközökbe sorolás, a gyakoriságok és relatív gyakoriságok kiszámítása

Osztályközökbe sorolás alatt a következőket értjük (a statisztikai irodalomban az osztálybasorolás kifejezés terjedt el, bár helyesebb lenne osztásközökbe sorolásról beszélni). Nagytömegű adat (sokelemű minta) esetén a statisztikai feldolgozás azzal kezdődik, hogy adatainkat nagyság szerint sorba rendezzük (rendezett minta). Megállapítjuk, hogy milyen intervallumba esnek az adatok és a teljes tartományon belül résztartományokat veszünk fel, azaz az eredeti intervallumot részekre osztjuk. Ezen részek neve osztályköz. Ezután megnézzük, hogy az egyes osztályközökbe hány adat került. Például, ha veszünk száz levegőmintát, az SO_2 koncentrációt meghatározzuk minden mintában és a legkisebb koncentráció 21 ppm, a legnagyobb 70, akkor meg lehet vizsgálni, hogy hány minta koncentrációja esett például a 21—30, 31—40, 41—50, 51—60, 61—70 ppm-es tartományokba. *Gyakoriságon* az egyes osztályközökbe eső elemek számát értjük. Jele f_i, ahol i az osztályköz sorszámát jelenti. Ha a gyakoriságot osztjuk a minta teljes elemszámával (esetünkben 120-szal), akkor a *relatív gyakoriságot* kapjuk. Jele: g_i.

Az előbbi példánkban legyenek adataink olyanok, hogy f_i értékeiként a táblázatban közölt adatokat kapjuk. Ekkor g_i-k (az f_i gyakoriságokhoz tartozó relatív gyakoriságok) éppen a következő táblázatban megadott értékűek lesznek.

	Osztályköz	f_i	g_i
1.	21—30	12	0,10
2.	31—40	24	0,20
3.	41—50	49	0,41
4.	51—60	23	0,19
5.	61—70	12	0,10

Tapasztalati eloszlásfüggvény és hisztogram meghatározása

Az eloszlásfüggvény tapasztalati úton becsülhető, a minta adatainak elemzése alapján. Tegyük fel, hogy a kísérletben meghatároztuk az x valószínűségi változó értékeit:

$$x_1, x_2, \ldots, x_n.$$

Ha kiszámítjuk az $F_t(x)$ értékeket (a t a tapasztalati szó jelölése) a következő képletből:

$$F_t(x) = \frac{1}{n} \sum_{x_i < x} 1,$$

akkor megkapjuk az empirikus eloszlási görbe ordinátaértékeit az adott x értékhez. A képlet szerint tehát $F_t(x)$ értéke az x-nél kisebb x_i értékű eredmények darabszá-

mának és a kísérletek számának hányadosa. $F_t(x)$ monoton növekvő (nem szigorú értelemben) és lépcsős.

Természetesen minden folytonos elméleti eloszlásfüggvényhez tartozik sűrűségfüggvény. Tegyük fel, hogy adatok eloszlása jól közelíthető valamilyen folytonos eloszlással. Folytonos eloszlás esetén, elég kis x-et választva, azt kapjuk, hogy

$$P(x \leqq \xi < (x+\Delta x)) = \int_{x}^{x+\Delta x} f(t)\,dt \cong f_t(x)\Delta x,$$

ahol $f(x)$ a tapasztalati sűrűségfüggvény, azaz a sűrűségfüggvény tapasztalati értéke. Ez a formula lehetővé teszi a sűrűségfüggvény empirikus közelítését. Osszuk fel azt az intervallumot, amelybe ξ lehetséges értékei esnek sok kis, egyenként Δx hosszúságú intervallum összegére. A kísérletet (amellyel ξ kapcsolatos), végezzük el n-szer és jegyezzük fel a ξ-re kapott értékeket. Végezzük el az osztályközökbe sorolást és számítsuk ki a relatív gyakoriságokat. Az egyes osztályközök relatív gyakoriságának megfelelő magasságú téglalapot emelve az osztályköz abszcissza intervalluma fölé, és ezt minden osztályközre elvégezve kapjuk az ún. *hisztogramot.* Ez is lépcsős függvény. A hisztogramok vizsgálata az eloszlás jellegének megállapításánál jó szolgálatot tehet.

A legnagyobb gyakoriságú osztályköznek külön nevet is adtak: ez a modusz. Ez a fogalom már átvezet a minta jellemző értékeinek számításához. Előbb azonban nézzük meg, mit is jelent a minta osztályba sorolása a gyakorlatban.

Legyen feladatunk egy 50 elemű minta feldolgozása. Az adatok az élet bármely területeiről származhatnak: beszélhetnénk 50 tehén napi tejhozamáról, 50 felnőtt férfi vérének koleszterinszintjéről, egy nagyváros egy forgalmas pontján 50 egymást követő napon a levegő kén-dioxid tartalmának átlagértékéről stb.

Legyenek adataink a következő számértékekkel jellemezve:

Sorszám (n)	Érték (x)	(n)	(x)	(n)	(x)	(n)	(x)	(n)	(x)
1	10	11	12	21	10	31	6	41	21
2	12	12	11	22	15	32	16	42	12
3	11	13	13	23	16	33	13	43	16
4	18	14	13	24	16	34	16,5	44	15
5	16	15	14	25	15,5	35	15	45	16
6	17	16	15	26	12	36	15	46	17
7	15	17	11	27	13	37	14	47	16
8	10	18	16	28	15	38	13	48	11
9	9	19	17	29	10	39	12	49	10
10	20	20	17	30	8	40	10	50	8

Adatainkat kényelmesen feldolgozhatjuk az ALAPVETŐ STATISZTIKAI JELLEMZŐK menü MINTAFELDOLGOZÁS OSZTÁLYBA SOROLÁSSAL programjával (2. menüág).

A program először megkérdezi, hogy adataink lemezen vannak vagy a billentyűzetről kívánjuk azokat bevinni. Tegyük fel, hogy első alkalommal billentyűzetről visszük be adatainkat. Először természetesen az adatok számát (a minta elemeinek számát) kéri a program, majd egyenként a minta elemeit.

A program a feldolgozáskor először átnézi az adatokat. Kiírja a minta legkisebb és legnagyobb elemét. Ez tájékoztatás a felhasználónak, segít az osztályhatárok kijelölésében. Ezután minden osztályköz felső osztályhatárát kell megadni. Legyenek most az osztályok 3 egységnyi szélességűek, azaz a felső osztályhatárok (mivel

a legkisebb elem 6, a legnagyobb 21) rendre 9, 12, 15, 18, 21. Ennek megadása után a program kiírja (az adott példa esetén): VAN OLYAN OSZTÁLYKÖZ, AMELYBEN 8-NÁL KEVESEBB ADAT VAN. (Ez nem hibajelzés, csak tájékoztatás. Ezen a ponton a felhasználó megváltoztathatja az osztályhatárokat, hogy megfelelően nagy számú elem essék minden osztályba.)

Ha az osztályba sorolást helybenhagytuk, megjelenik a következő almenü a képernyőn:

AZ ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:

GYAKORISÁGI TÁBLÁZAT	(A)
MODUSZ KISZÁMÍTÁSA	(B)
TAPASZTALATI ELOSZLÁSFÜGGVÉNY KIRAJZOLÁSA	(C)
HISZTOGRAM RAJZOLÁSA	(D)
VÉGE	(E)

Tegyük fel, hogy *minden* lehetséges számolást szeretnénk elvégezni mintáinkkal. Válasszuk először az (A) lehetőséget. A képernyőre kiíródik a következő táblázat:

GYAKORISÁGTÁBLÁZAT

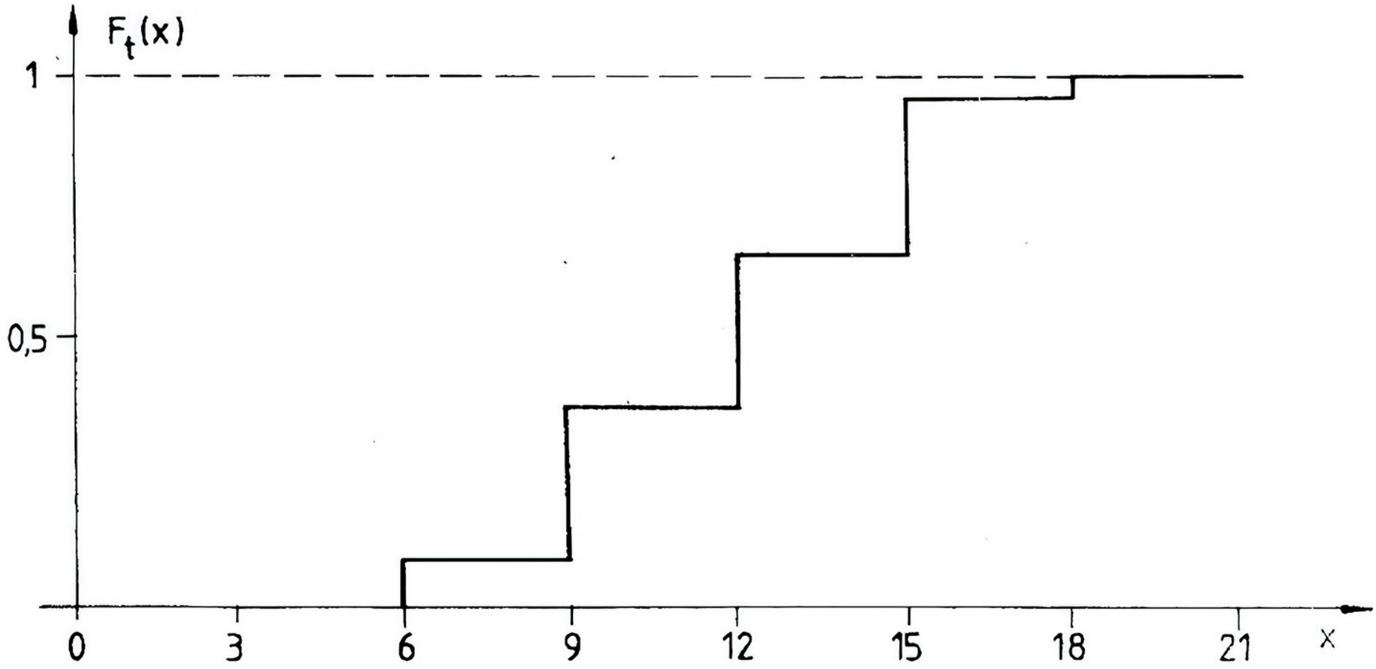
OSZTÁLYKÖZ	GYAKORISÁG	RELATÍV GYAKORISÁG
(6)< 9	4	0,08
<12	15	0,3
<15	14	0,28
<18	15	0,3
<21	2	0,04

A szóköz billentyű megnyomása után felvehetjük eredményeinket mágneslemezre, de enélkül is tovább számolhatunk a bevitt adathalmazzal. Bármely billentyű megnyomása után tovább számolhatunk a programmal.

A (B) lehetőséget választva a moduszt írathatjuk ki a képernyőre (vagy nyomtatóra). Esetünkben a kiírt eredmény:

NINCS EGYÉRTELMŰ MEGOLDÁS, TÖBB AZONOS GYAKORISÁGÚ OSZTÁLYKÖZ

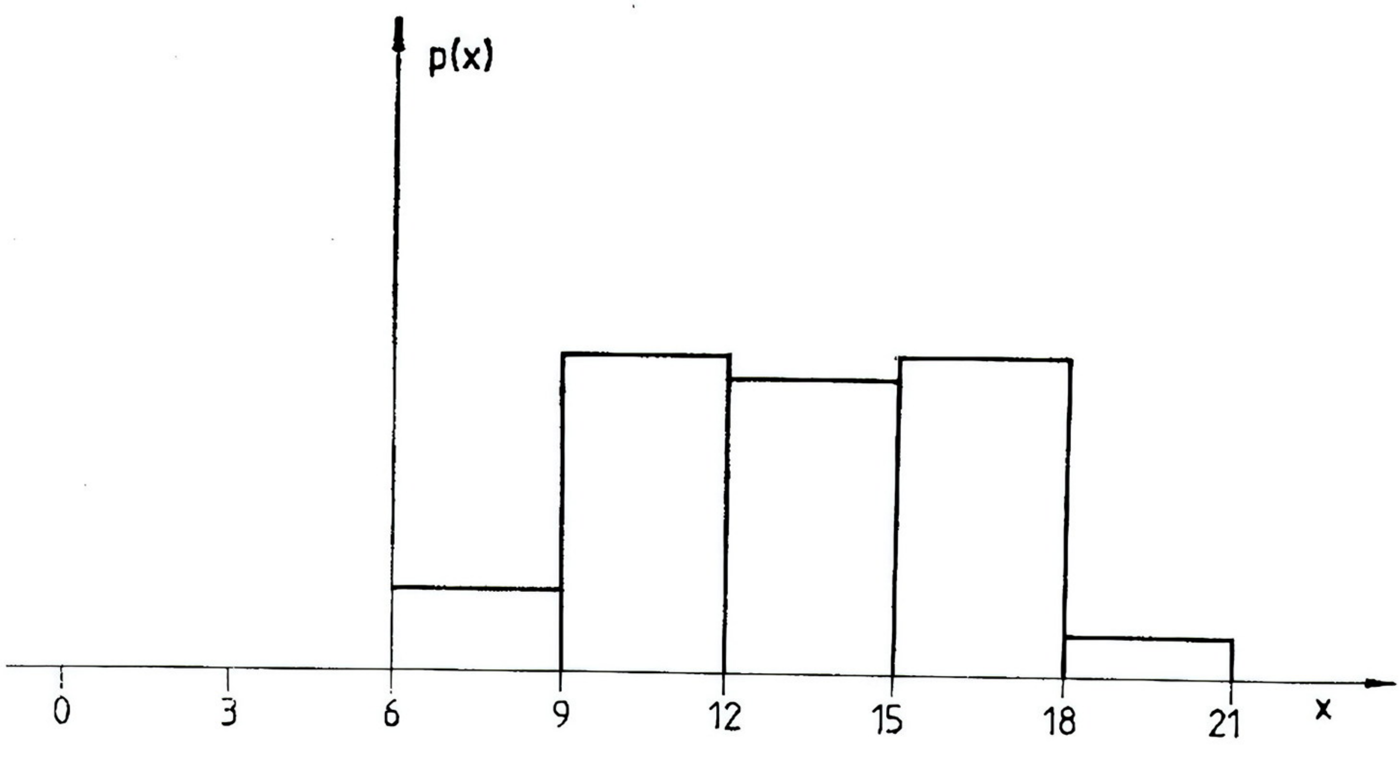
Az eredmény érthető, hiszen az előbb láttuk, hogy mind a $9 \leqq x < 12$, mind a $15 \leqq x < 18$ osztályközbe 15—15 adat esik. Rajzoljuk ezután ki a képernyőre a tapasztalati eloszlásfüggvényt a mintára (C). A következő kép jelenik meg:



11. ábra. Tapasztalati eloszlásfüggvény

Az eloszlásfüggvény képét a nyomtatóval is kirajzoltathatjuk.

Ezután a (D) lehetőséget választva a menüből, a hisztogram jelenik meg a képernyőn, és a kapott ábrát igény szerint ismét kinyomtathatjuk.



12. ábra. Hisztogram

MINTAFELDOLGOZÁS OSZTÁLYBA SOROLÁSSAL

A változók listája:

N	— a mintacsoport elemszáma,
X(I)	— a mintacsoport elemei,
W	— segédváltozó,
D	— az osztályközök száma,
L(I)	— az osztályközök felső határai,
K(J)	— a gyakoriságok összege a j-edik osztályközig,
P(J), R(J)	— a gyakoriságok,
OX	— adatmátrixból kijelölt oszlop száma,
A$	— igen-nem választás,
C$	— menünél választás,
ID	— kiíratásnál választás (képernyő vagy nyomtató),
N$	— adatfile neve,
HS, HN$, S, SZ	— lemezhiba ellenőrzése,

Programmagyarázat:

100:	Programcsomag-szervezés.
110—250:	A program neve.
260—1200:	Adatbevitel, számítások
1210—1370:	Menü kiírása, választás.
1380—1760:	Eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
1770:	Visszatérés a FŐMENÜ c. programra
1780—1900:	A program által nyújtott lehetőségek rövid ismertetése.
1910—2260:	Adatbeolvasás lemezről. Közben lemezhiba ellenőrzése a 2270-es szubrutinnal.
2270—2380:	Lemezhibát ellenőrző szubrutin.
2390—2490:	Várakozó szubrutin.
2500—2570:	Várakozó szubrutin Igen-Nem válaszra.
2580—2920:	Grafikonok felrajzolása.

```
500 REM*******************************
510 REM*                             *
520 REM* AZ ADATOK SORBARENDEZESE    *
530 REM*                             *
540 REM*******************************
550 PRINT CHR$(147):PRINT"SORBARENDEZES"
560 FOR I=1 TO N-1:CS=0:PRINT"SQQ";I
570 FOR J=1 TO N-I:PRINT"SQQQ";J
580 IF X(J)>X(J+1) THEN W=X(J):X(J)=X(J+1):X(J+1)=W:CS=1
590 NEXT J:IF CS=0 THEN I=N
600 NEXT I
610 REM*******************************
620 REM*                             *
630 REM* OSZTALYKOZOKBE SOROLAS      *
640 REM*                             *
650 REM*******************************
660 PRINT:PRINT"A MINTACSOPORT LEGKISEBB ELEME :":PRINT X(1)
670 PRINT:PRINT"A MINTACSOPORT LEGNAGYOBB ELEME :":PRINT X(N)
680 IF X(1)<>X(N) THEN 710
690 PRINT:PRINT"MINIMUM=MAXIMUM, NINCS ERTELME AZ OSZTALYBA SOROLASNAK."
700 GOSUB 2450:GOTO 260
710 PRINT:PRINT"HANY OSZTALYKOZBE KIVANJA SOROLNI AZ     ADATOKAT?"
720 INPUT D
730 IF D>10 OR D<2 OR D<>INT(D) THEN PRINT"2 ES 10 KOZTI EGESZ!":GOTO 720
740 L(0)=X(1):L(D)=X(N)
750 HS=0:FOR I=1 TO D-1
760 PRINT"ADJA MEG A";I;".OSZTALYKOZ FELSO HATARAT!":INPUT L(I)
770 IF L(I)<=L(I-1) OR L(I)>L=(D) THEN PRINT"TEVES ADAT!":HS=1:I=D:GOTO 810
780 IF I<3 THEN 810
790 IF ABS((L(I)-L(I-1))/(L(I-1)-L(I-2))-1)<1E-3 THEN 810
800 PRINT"EGYENLO OSZTALYKOZOKET ADJON MEG!":I=D:HS=1
810 NEXT I:IF HS<>0 THEN 750
820 K(0)=0:K(D)=N
830 FOR J=1 TO D-1:FOR I=K(J-1)+1 TO N
840 IF X(I)>L(J) THEN K(J)=I-1:I=N
850 NEXT I:NEXT J
860 H=0:FOR I=1 TO D:P(I)=K(I)-K(I-1):IF P(I)<8 THEN H=1
870 NEXT I
880 IF H=0 THEN 910
890 PRINT"VAN OLYAN OSZTALYKOZ, AHOL 8-NAL          KEVESEBB ELEM VAN."
900 PRINT"KIVAN JAVITANI";:GOSUB 2550:IF A$="I" THEN 710
910 FOR J=1 TO D:R(J)=P(J):NEXT J
920 FOR I=1 TO D-1:H=0:FOR J=1 TO D-I
930 IF R(J)>R(J+1) THEN W=R(J):R(J)=R(J+1):R(J+1)=W:H=1
940 NEXT J:IF H=0 THEN I=D
950 NEXT I:GOTO 1010
```

```
2580 REM*********************************
2590 REM*                               *
2600 REM*  GRAFIKONOK FELRAJZOLASA      *
2610 REM*                               *
2620 REM*********************************
2630 SYS 3078:SYS 3072
2640 SYS 3084,-1,D+1,-.1,1.1
2650 SYS 3099,-1,0,D+1,0
2660 SYS 3099,0,-.1,0,1.05
2670 SYS 3099,-.2,.5,.2,.5
2680 SYS 3090,-1,.5,0,STR$(.5)
2690 SYS 3099,-.2,1,.2,1
2700 SYS 3090,-1,1,0,STR$(1)
2710 FOR I=0 TO D
2720 SYS 3099,I,.02,I,-.02
2730 NEXT I
2740 IF C$="D" THEN 2820
2750 FOR I=0 TO D-1
2760 SYS 3099,I,K(I)/N,I,K(I+1)/N
2770 SYS 3099,I,K(I+1)/N,I+1,K(I+1)/N
2780 NEXT I
2790 SYS 3090,1,1,0,"TAPASZTALATI"
2800 SYS 3090,1,.9,0,"ELOSZLASFGV."
2810 GOTO 2880
2820 FOR I=0 TO D-1
2830 SYS 3099,I,0,I,P(I+1)/N
2840 SYS 3099,I,P(I+1)/N,I+1,P(I+1)/N
2850 SYS 3099,I+1,P(I+1)/N,I+1,0
2860 NEXT I
2870 SYS 3090,1,1,0,"HISZTOGRAM"
2880 LE$="[illegible]"
2890 PRINT LE$;"KIRAJZOLJUK PRINTERRE IS";:GOSUB 2550
2900 IF A$="I" THEN SYS 3087
2910 SYS 3075
2920 RETURN
```

3.3. A minta jellemzése (középértékek, tapasztalati szórás)

Kétféle középértéktípust különböztetünk meg. A *számított középértékek* vagy *átlagok* számítások (aritmetikai műveletek) eredményei, így a minta elemeivel matematikai összefüggésben állnak, és értékük az adatok sorrendjétől független. A *helyzeti középértékek* az értékek nagysága szerint rendezett mintában általában matematikai műveletek nélkül kijelölhetők.

a) Helyzeti középértékek. Ilyen a medián és a modusz. A *medián* (ME) az a középérték, amely az x_i mennyiségek értékeinek sokaságát 2 egyenlő számú tagból álló részre osztja úgy, hogy a sokaság egyik felében az x_i értékek a mediánnál kisebbek, a másikban nagyobbak. Páratlan elemszámú mintánál a középső érték a medián, páros elemszámúnál a két középső összegének fele. Információtartalma az eloszlás típusától függ, de torzítatlan. *Modusznak* nevezzük a véletlen mennyiség legnagyobb

gyakoriságú értékét vagy értékeit (osztályba sorolásnál a legnagyobb gyakoriságú osztályközt).

b) Számított középértékek (számtani átlag, négyzetes átlag, harmonikus átlag, logaritmikus átlag).

Legfontosabb a számtani átlag. Jele: $\bar{x}$.

Kiszámítása:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},$$

ahol x_i az i-mintaelem adott tulajdonságának számértéke; n a minta elemszáma, a feldolgozásba bevont mérések száma.

Statisztikai feldolgozásnál, ha adatainkat már osztályközökbe soroltuk, a számtani átlagot a gyakoriságokból vagy a relatív gyakoriságokból is kiszámíthatjuk:

$$\bar{x} = \frac{\sum_{k=1}^{r} f_k x_k}{\sum_{k=1}^{r} f_k} = \sum_{k=1}^{r} g_k x_k,$$

ahol r az osztályközök száma; k az illető osztályköz sorszáma; x_k az azonos osztályba eső elemek jellemzője, pl. az osztályközép a számtani átlaggal becsülve; f_k az x_k gyakorisága; g_k az x_k relatív gyakorisága.

A többi középértéktípus számunkra kevésbé fontos, de megemlítjük, hogy négyzetes középértékekkel akkor dolgozunk, ha valamelyik mennyiség egy másik változó négyzetétől függ. Harmonikus átlaggal akkor, ha a reciprokától, logaritmikus átlaggal akkor, ha a folyamat logaritmikus összefüggésekkel írható le stb.

A legfontosabb átlagképzési szabályok a következők.

Mértani (geometriai) átlag:

$$x_G = \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

Megjegyzés:

$$\prod_{i=1}^{n} x_i = x_1, x_2, \ldots, x_n.$$

Harmonikus átlag:

$$\bar{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}.$$

Négyzetes átlag:

$$\bar{x}_N = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}.$$

Logaritmikus átlag:

$$\bar{x}_L = e^{\frac{\sum_{i=1}^{n} \log(x_i)}{n}}.$$

A középértékek semmit sem mondanak az adatok szóródásáról, hiszen ugyanaz a középérték csupa azonos értékű mintaelem alapján is, de egészen különbözőekből is kijöhet. Ezért szükség van a szóródás jellemzésére is.

Az adatok szóródását jellemezhetjük a minta terjedelmével. Ez a minta legnagyobb és legkisebb elemének különbsége. Hasonlóan a moduszhoz és a mediánhoz, a rendezett mintában további számítások nélkül, közvetlenül megadható. A minta elemeinek eloszlásáról azonban keveset mond. A minta elemzésénél a minta szóródását a korrigált tapasztalati szórásnégyzettel szokás jellemezni:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}.$$

Bizonyítható, hogy a korrigált tapasztalati szórásnégyzet várható értéke a valószínűségi változó elméleti szórásnégyzete, így az elméleti szórásnégyzet becslésénél a korrigált tapasztalati szórásnégyzetet kell használni.

Megjegyezzük, hogy a régebbi szakirodalomban gyakran a tapasztalati szórásnégyzet alatt korrigálatlan értékét értik, azaz az előbbi képlet nevezőjében $n-1$ helyett n áll:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}.$$

Ha az $x_i - \bar{x}$ értékeket vektor alakba rendezzük, pl. a következő módon:

$$\begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} = d,$$

akkor a korrigált tapasztalati szórásnégyzetet fejezik ki a következő képletek is:

$$s^2 = \frac{\sum d_i^2}{n-1}, \quad s^2 = \frac{\sum_{i=1}^{r} f_k d_k^2}{\sum_{k=1}^{r} f_k - 1},$$

ahol $d_k = x_k - \bar{x}$.

Az utolsó képletben r az osztályközök száma, d_k az osztályköz átlag és a teljes minta átlagának eltérése.

Az ALAPVETŐ STATISZTIKAI JELLEMZŐK menüjéből az ÁTLAG-, MEDIÁN- ÉS SZÓRÁSSZÁMÍTÁS (1) lehetőséget választva lehetőségünk van különböző átlagolási szabályok szerinti átlagképzési módokat kipróbálni, akár ugyanarra az adatsorra. Gyakorlásképpen tekintsük a következő 10 adatot: 10, 12, 11, 18, 16, 17, 15, 10, 9, 20. A megfelelő menüág kiválasztása után természetesen ez alkalommal is lehetőségünk van rövid ismertetőt kérni a programról, majd dönthetünk arról, hogy lemezen levő adatokat kívánunk feldolgozni vagy adatainkat a billentyűzetről visszük be. Tekintsük az egyszerűség kedvéért az utóbbit. Mintánk 10 elemű, ezért A MINTA ELEMEINEK SZÁMA: 10, a 10-es szám begépelése után be kell írni a minta elemeit. Ezután megjelennek az összes említett értékek:

SZÁMTANI ÁTLAG	13,8
GEOMETRIAI ÁTLAG	13,31
HARMONIKUS ÁTLAG	12,84
NÉGYZETES ÁTLAG	14,28
LOGARITMIKUS ÁTLAG	13,31
MEDIÁN	13,5
SZÓRÁSNÉGYZET	15,07
SZÓRÁS	3,88

Látható a táblázat adataiból, hogy a különféle átlagképzési szabályok eltérő átlagértékeket adnak. Az eltérés mértéke az adatok eloszlásától függ. Felhívjuk a figyelmet arra, hogy ha a sokaság, amiből a minta származik, normális eloszlású, akkor a számtani átlag a várható érték legjobb becslésének tekinthető, és ez esetben a korrigált tapasztalati szórás az eloszlás szórását jól becsüli.

ALAPVETŐ STATISZTIKAI JELLEMZŐK

A változók listája:

N	— a mintacsoport elemszáma,
X(I)	— a mintacsoport elemei,
M1	— számtani átlag,
M2	— geometriai átlag,
M3	— harmonikus átlag,
M4	— négyzetes átlag,
M5	— logaritmikus átlag,
MM	— medián,
Q	— korrigált tapasztalati szórásnégyzet,
Z(I, J)	— lemezről beolvasott adatok,
OX	— adatmátrixból kijelölt oszlop száma,
A$	— Igen-Nem választás,
HS, HN$, S, SZ	— lemezhiba-ellenőrzés.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—230:	A program neve.
240—1020:	Adatbevitel, számítások.
1030—1240:	Az eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
1250:	Programcsomag-szervezés.
1260—1620:	Adatbeolvasás lemezről.
1630—1740:	Lemezhiba-ellenőrzés.
1750—1870:	A program által nyújtott lehetőségek rövid ismertetése.
1880—2060:	Várakozás Igen-Nem válaszra.

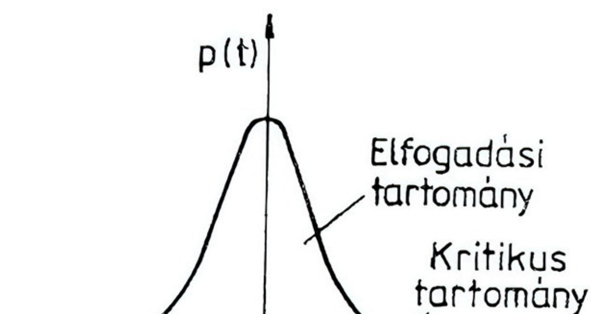
```
700 REM*****************************
710 REM*                           *
720 REM* SZAMITASOK                *
730 REM*                           *
740 REM*****************************
750 M1=0:M2=1:M3=0:M4=0:M5=0:Q=0
760 E2=0:E3=0:E5=0
770 FOR I=1 TO N:M1=M1+X(I):NEXT I:M1=M1/N
780 FOR I=1 TO N
790 Q=Q+(X(I)-M1)*(X(I)-M1)
800 IF E2=1 THEN 830
810 IF X(I)<0 THEN E2=1:GOTO 830
820 M2=M2*X(I)↑(1/N)
830 IF E3=1 THEN 860
840 IF X(I)=0 THEN E3=1:GOTO 860
850 M3=M3+1/X(I)
860 M4=M4+X(I)*X(I)
870 IF E5=1 THEN 900
880 IF X(I)<=0 THEN E5=1:GOTO 900
890 M5=M5+LOG(X(I))
900 NEXT I
910 FOR I=1 TO N-1
920 FOR J=I+1 TO N
930 IF X(I)>X(J) THEN W=X(I):X(I)=X(J):X(J)=W
940 NEXT J:NEXT I
950 Q=Q/(N-1):IF M3=0 THEN E3=1
960 IF E3=0 THEN M3=1/M3
970 M4=SQR(M4/N):IF E5=1 THEN 1000
980 M5=M5/N:IF M5>LOG(1E38) THEN E5=1:GOTO 1000
990 M5=EXP(M5)
1000 IF INT(N/2)=N/2 THEN MM=(X(N/2)+X(N/2+1))/2:GOTO 1020
1010 MM=X((N+1)/2)
1020 PRINT
```

4. STATISZTIKAI PRÓBÁK

4.1. A statisztikai próbákról általában

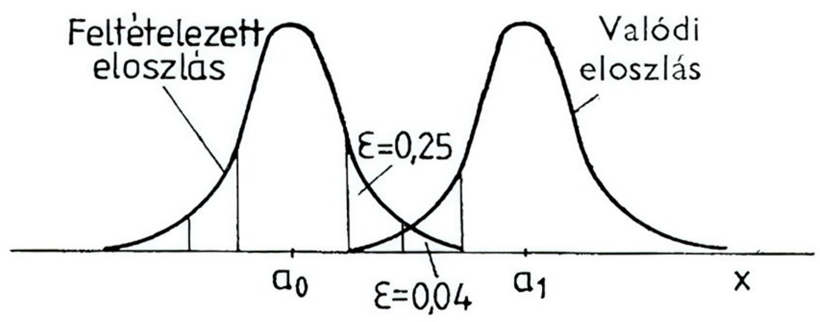
A statisztikai próbák egy vagy több statisztikai sokaság eloszlásaival kapcsolatos hipotézisek ellenőrzésére alkalmasak. A hipotézis vonatkozhat az eloszlás típusára, vagy az eloszlás típusának ismerete esetén az eloszlás egyes paramétereire. A hipotézisvizsgálatok közös vonása az, hogy a nullhipotézist igaznak tételezzük fel, majd ezen feltételezés mellett megállapítjuk, hogy a próbastatisztikák mint valószínűségi változók, milyen eloszlásúak. A valószínűségi változó szélső értékeit, amelyek bekövetkezésének már csekély a valószínűsége, kritikus tartománynak tekintjük. Ha egy konkrét próbában a próbastatisztika értéke ebbe az ún. kritikus tartományba esik, a nullhipotézist elvetjük a rendelkezésre álló minta alapján. A minta elemzése és a statisztikai próba tehát azt a kockázatot foglalja magában, hogy esetleg az igaz hipotézist elvetjük, ill. a téveset elfogadjuk.

A gyakorlatban a kritikus tartomány kijelölése úgy történik, hogy az eloszlás egyik vagy mindkét végéből levágunk adott valószínűségi szinthez egy-egy darabot (13. ábra) és azt mondjuk, hogy ami oda esik, az már nem tartozik a nullhipotézisben feltételezett eloszláshoz.



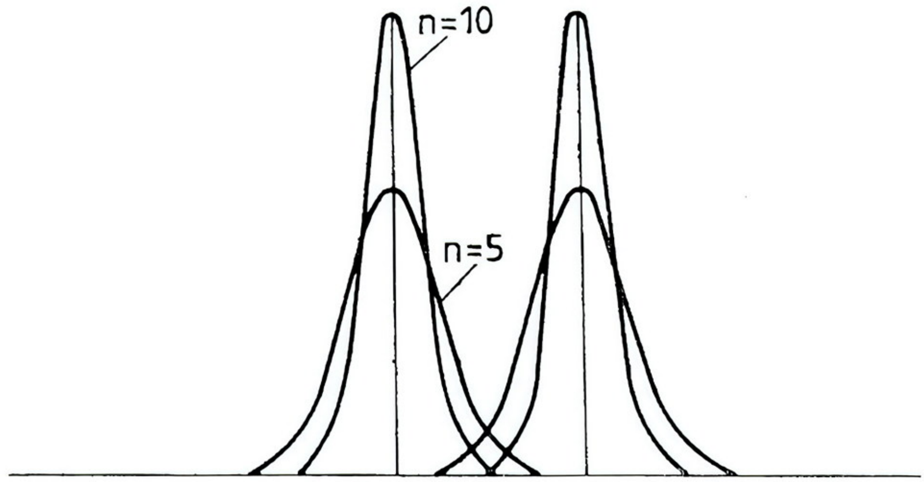
13. ábra Kritikus tartomány kijelölése

Mennyit vágjunk le? Mindenesetre annyit, hogy az igaz hipotézis elvetésének valószínűsége ne legyen túl nagy. Miért nem vállalunk egészen kicsi kockázatot? Az 5%-os kockázatvállalás pl. azt jelenti, hogy az esetek 5%-ában, tehát nagyjából minden huszadik esetben tesszük ki magunkat annak a tévedésnek, hogy az egyébként igaz nullhipotézist elvetjük. A valószínűségi szintet csökkentve, az ilyen irányú hiba csökken, de fellép egy ellenkező irányú hiba: annak lehetősége, hogy a vizsgált paraméter nem a nullhipotézisben feltételezett eloszlásból való, de kiválasztott értéke „véletlenül" a nullhipotézisben feltételezett eloszlás elfogadási tartományába esik. Ezt szemlélteti a 14. ábra.



14. ábra. Kritikus tartomány elhelyezkedése két átlapoló eloszlás feltételezésével

Ez ellen a hiba ellen elsősorban úgy védekezhetünk, ha a minta elemszámát növeljük. A szórás az elemszám négyzetgyökével fordítva arányos. Nagyobb elemszámú minta esetén tehát a szóródás csökken, és a sűrűségfüggvények „karcsúbbak" lesznek, így nem nyúlnak annyira egymásba. Ezt szemlélteti a 15. ábra.



15. ábra. A minta elemszámának hatása a minta eloszlására

Ezzel kapcsolatban ki kell térnünk a szignifikancia fogalmára. A nullhipotézis elvetéséhez általában az kell, hogy a vizsgált statisztikai jellemző nullhipotézisben feltételezett és valódi értéke elég messze legyen egymástól. Ilyen esetben (a nullhipotézis elvetésekor) azt mondjuk, hogy az eltérés *szignifikáns*. Ez azt jelenti, hogy annak a valószínűsége, hogy az eltérést a véletlen okozza, nagyon kicsi. A szignifikanciaszint az eloszlás „levágott" darabja alatti terület mértéke. 5%-os szignifikanciaszint azt jelenti tehát, hogy az esetek *legfeljebb* 5%-ában lehetséges, hogy az eltérés a nullhipotézisben feltételezett értéktől csak a véletlen hatások következménye. Mivel a 10%-os szignifikanciaszint esetén ez a tévedés az esetek 10%-ában is megengedett, a 10%-os szignifikanciaszint ebben az értelemben „gyengébb".

A statisztikai próbákat többféle szempont szerint osztályozhatjuk. Ha azt vizsgáljuk, hogy egy adott minta normális eloszlású sokaságból származik-e, akkor ún. normalitásvizsgálatot végzünk. Ha a valószínűségi változónkról tudjuk, hogy normális eloszlást követ, akkor a normális eloszlás két paramétere, a várható érték

és a szórás feltételezett értékének ellenőrzése lehet a feladat. Ha a sokaság elméleti szórásnégyzete ismert, a várható értékre vonatkozó hipotézist *egymintás u-próbával,* ha a szórást is a minta alapján kell becsülnünk, *egymintás t-próbával* ellenőrizhetjük, az utóbbi esetben azzal a feltételezéssel, hogy a minta alapján számított tapasztalati szórás várható értékben megegyezik a sokaság szórásával.

Vizsgálható két paraméter egyezése is statisztikai próbákkal. Normális eloszlású valószínűségi változók esetén, ha a várható értékek egyezésére vonatkozó hipotézist vizsgáljuk, és az elméleti szórások ismertek és azonosak, akkor az ún. *2-mintás u-próbát* alkalmazhatjuk. Ha az elméleti szórásnégyzeteket a minta tapasztalati szórásnégyzeteivel becsüljük, akkor az ún. *2-mintás t-próbával* dolgozhatunk; azzal a feltételezéssel, hogy az ismeretlen elméleti szórások megegyeznek. Két szórás egyezésének vizsgálatára alkalmas az *F-próba.* Ha az elméleti szórások nem ismertek, azaz értéküket a korrigált tapasztalati szórásokkal becsüljük, s az F-próba szerint nem egyeznek meg egymással, akkor a várható értékek egyezésének vizsgálatára a *Welch-próba* használható. Több szórás egyezésének egyidejű vizsgálatát a *Bartlett-próbával* végezhetjük.

4.2. Függetlenségvizsgálat χ^2 próbával

A gyakorlatban gyakran felmerül az a kérdés, hogy egy véletlen esemény kimenetelét valamilyen tényező befolyásolja-e vagy sem. Ilyen befolyásoló tényező (faktor) lehet pl. a hely, idő, valamilyen mesterséges behatás (kezelés) stb. Tekintsük a következő példát: Egy gyár két különböző üzeme gyárt porszívót. Egy-egy szállítmányt kiválasztva mindkét üzemből, 6 hónapon át figyelemmel kísérik az eladott porszívók meghibásodását, azaz azt, hogy az első 6 hónapban hány darabot hoztak vissza garanciális javításra. A vizsgálat eredményéből arra szeretnének következtetni, hogy vajon az egyik gyár szignifikánsan jobb terméket (kevesebb garanciális javítást igénylő terméket) gyárt-e, mint a másik vagy sem. A kérdésfelvetés egyenértékű azzal, mintha azt kérdeznénk, hogy a 6 hónapon belüli meghibásodás független-e attól, hogy melyik üzemben gyártották a porszívót.

Az egyes üzemekből kibocsátott porszívók meghibásodási statisztikája esetünkben legyen a következő:

	Hibátlan működésű (db)	Garanciális javításra küldött (db)
A üzem	10	31
B üzem	23	48

A táblázatban voltaképpen a porszívókat két jellemzőjük szerint osztályoztuk: a gyártás helye szerint és a 6 hónapon belüli meghibásodás bekövetkezése szerint. Ennek alapján kétszeres osztályozást végeztünk. Az eredményt ún. kétmezős kontingenciatáblázatban foglaltuk össze. Általánosan a kontingenciatáblázatnak legyen r sora és c oszlopa. Végezzünk n véletlen kísérletet, és jelölje n_{ij} azon kísérletek számát, amelyek az i sor j oszlopába esnek. Így a következő táblázatot kapjuk:

$$\begin{array}{cccc} n_{11} & n_{12} \ldots n_{1c} & n_{1.} \\ n_{21} & n_{22} \ldots n_{2c} & n_{2.} \\ \vdots & & \vdots \\ n_{r1} & n_{r2} \ldots n_{rc} & n_{r.} \\ \hline n_{.1} & n_{.2} \ldots n_{.c} & n \end{array}$$

ahol $n_{i\cdot} = \sum_j n_{ij}$ az i sorba eső megfigyelések száma, és $n_{j\cdot} = \sum_i n_{ij}$ a j oszlopba eső megfigyelések száma.

Általánosan fogalmazva arra vagyunk kíváncsiak, hogy a kétféle osztályozás egymástól független-e. Az előző példánál maradva, ha az A üzem nagyobb valószínűséggel bocsát ki 6 hónapon belül hibátlanul működő porszívókat, mint a B üzem, akkor a kibocsátás helyétől nem független a meghibásodás valószínűsége. Ezért nullhipotézisünket a következőképpen fogalmazhatjuk meg: Legyen p_{ij} annak valószínűsége, hogy egy véletlen kísérlet eredménye az i. sor j. oszlopába esik. Legyen $p_{i.}$ annak a valószínűsége, hogy a véletlen megfigyelés az i. sorba esik, $p_{.j}$ pedig annak a valószínűsége, hogy a j. oszlopba. Nullhipotézisünk az, hogy a sorok és oszlopok egymástól függetlenek:

$$H_0\colon\ p_{ij} = p_{i.} \cdot p_{.j} \quad \text{minden } i, j\text{-re.}$$

A $p_{i.}$ és $p_{j.}$ valószínűségeket az illető esemény relatív gyakoriságával becsülhetjük:

$$p_{i.} \approx \frac{n_{i.}}{n} \quad \text{és} \quad p_{.j} \approx \frac{n_{.j}}{n}.$$

Ezért, ha H_0 igaz, az i. sor j. oszlopába esés valószínűségét a következőképpen becsülhetjük:

$$np_{ij} = n.p_{i.}.p_{.j} \approx n\frac{n_{i.}}{n}\frac{n_{.j}}{n} = \frac{n_{i.}.n_{.j}}{n}.$$

Ezt hasonlíthatjuk a megfigyelt n_{ij} gyakorisághoz, és így a következő próbastatisztikát alakíthatjuk ki:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left\{\frac{\left(n_{ij} - \frac{n_{i.}.n_{.j}}{n}\right)^2}{n_{i.}.n_{.j}/n}\right\}.$$

Levezethető, hogy a fenti próbastatisztika szabadsági foka

$$(r-1)(c-1).$$

Ha a nullhipotézis igaz, a próbastatisztika χ^2 eloszlást követ $(r-1)(c-1)$ szabadsági fokkal. Ha a kiszámított érték nagyobb, mint a χ^2 eloszlás kritikus értéke a kiválasztott szignifikanciaszinten, a nullhipotézist, azaz a kétféle osztályozás függetlenségét el kell vetnünk. A fejezet bevezetőjében feltett kérdés megválaszolására hívjuk be a STATISZTIKAI PRÓBÁK menüjéből a FÜGGETLENSÉGVIZSGÁLAT χ^2 PRÓBÁVAL programot. A képernyőn megjelenő kérdésekre értelemszerűen válaszolva, az ismertetés elolvasása után, vagy anélkül eljutunk az adatok megadásának fázisába. Az egyszerűség kedvéért billentyűzetről vigyük be az adatokat.

A MINTACSOPORTOK SZÁMA:
(A TÁBLÁZAT SORAINAK SZÁMA)

Esetünkben két minta van, az A, ill. B üzemben gyártott porszívóké, a válasz 2.

A KÜLÖNBÖZŐ ESEMÉNYEK SZÁMA:
(A TÁBLÁZAT OSZLOPAINAK SZÁMA)

Mivel az események itt a 6 hónapon belüli javítás bekövetkezése, ill. be nem következése, két (komplementer) eseményünk van, ezért a válasz 2.

Ezután meg kell adni a kontingenciatáblázat adatait soronként. A kontingenciatáblázat adatainak begépelése után szinte azonnal kiíratható a képernyőre (vagy nyomtatóra) az eredmény:

VALAMENNYI REKESZRE (CELLÁRA)

ÉSZLELT ÉRTÉK	ELVÁRT ÉRTÉK	KHI-NÉGYZET
	1. OSZLOP	
10	12,08	0,358
23	20,92	0,267
	2. OSZLOP	
31	28,91	0,150
48	50,08	0,086

KHI-NÉGYZET =0,8012
SZABADSÁGI FOK=1

Hogyan kell értelmezni az előbbi eredményeket?

Az eredmények első oszlopához nem szükséges magyarázatot fűzni. A második oszlopban álló számok azt mutatják, hogy amennyiben az illető esemény mindkét üzem termékei között egyforma valószínűséggel fordulna elő, akkor milyen előfordulási gyakoriságot várhatnánk. Például, ha azt az eseményt nézzük, hogy a 6 hónapos hibátlan működés az A üzem termékeinél hányszor fordul elő, akkor az azonos valószínűséggel való előfordulás azt jelenti, hogy a 41, A üzemből származó porszívónál ugyanazt a valószínűséget várjuk, mint az összes (A és B üzemből kikerült) porszívónál. Ezért az elvárt érték:

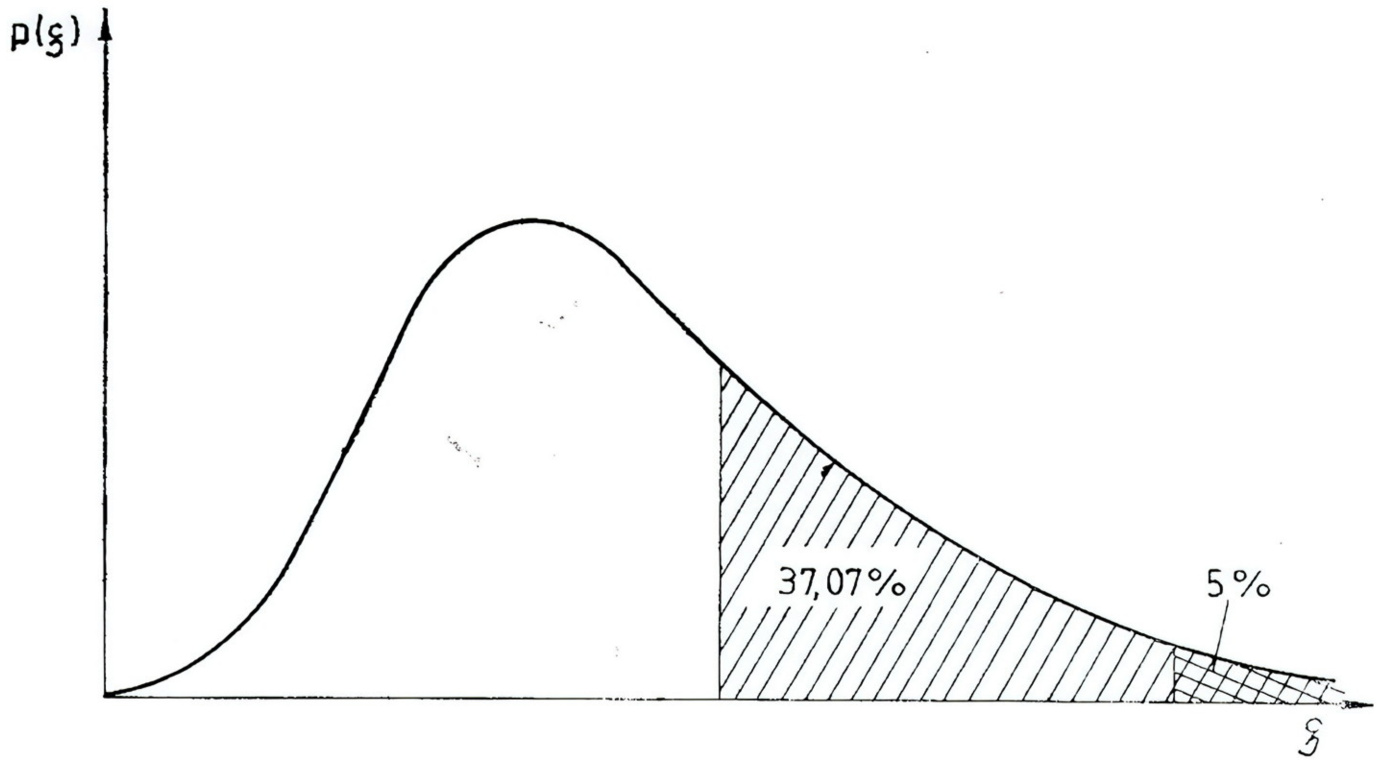
$$41 \cdot \frac{33}{112} = 12{,}08.$$

A harmadik oszlopban az illető eltérés hozzájárulása áll a kiszámított χ^2 értékhez (tudjuk, hogy a χ^2 érték additív). Az utolsó oszlop összege a keresett χ^2 érték.

Hogyan kell értelmezni a 0,8012-es kiírt χ^2 értéket? Először azt kellene megmondani, hogy a szóban forgó (1 szabadsági fokú) khi-négyzet eloszlásnál milyen valószínűséggel lesz a χ^2 eloszlású valószínűségi változó értéke kisebb, mint 0,812. Ha ez a valószínűség kisebb, mint az előre megadott 0,95-ös érték (95%-os megbízhatóságú döntés), az azt jelenti, hogy még ennél kisebb megbízhatósággal sem tudjuk azt mondani, hogy a két valószínűség különbözik, ezért a valószínűségek egyezését el kell fogadnunk.

A 0,812-es értékhez tartozó keresett valószínűséget az ELOSZLÁSOK menü KHI-NÉGYZET ELOSZLÁS programjával számíthatjuk ki (használatát lásd ott). A 0,8012-höz tartozó valószínűség: 0,6293. A választ úgy kell értelmezni, hogy $1-0{,}6293=0{,}3707$ a valószínűsége annak, hogy a két üzem azonos valószínűséggel

bocsát ki 6 hónapon keresztül meg nem hibásodó porszívót, és mi mégis különbözőnek tekintettük ezt a valószínűséget. Mivel azonban mi a kérdésre $1-0{,}95=0{,}05$ szignifikanciaszinten kívánunk válaszolni, azt kell mondanunk, hogy ekkora (37,07%-os) kockázatot az igaz hipotézis elvetésére nem vállalhatunk. Azt kell mondanunk tehát, hogy 5%-os szignifikanciaszinten a 6 hónapon belüli meghibásodások valószínűsége a két üzemből kikerülő porszívónál azonosnak mondható. Ezt szemlélteti a 16. ábra.



16. ábra. Illusztráció a példa megoldásához

FÜGGETLENSÉGVIZSGÁLAT KHI-NÉGYZET PRÓBÁVAL

A változók listája:

N	— a kontingencia táblázat sorainak száma,
M	— a kontingencia táblázat oszlopainak száma,
X(I, J)	— a kontingencia táblázat elemei,
I, R, K1	— segédváltozók,
G, T(I), V(I)	— összegzések, részeredmények,
E	— elvárt érték,
EX	— E kerekítve,
H	— khi-négyzet értéke egy adott értéknél,
HX	— H kerekítve,
C2	— khi-négyzet értéke összegezve,
M, N	— lemezről beolvasott adatmátrix oszlopainak és sorainak száma,
R	— adatmátrixból kijelölt oszlopok száma,
O(I)	— a vizsgált oszlopok,
A$	— Igen-Nem választás,
ID	— kiíratásnál választás (képernyő vagy nyomtató),
N$	— adatfile neve,
HS, HN$, S, SZ	— lemezhiba ellenőrzése,

Programmagyarázat:

100:	Programcsomag-szervezés.
110—240:	A program neve.
250—910:	Adatbevitel, számítások.
920—1250:	Eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
1260:	Visszatérés a FŐMENÜ c. programra.
1270—1500:	A program által nyújtott lehetőségek rövid ismertetése.
1510—1920:	Adatbeolvasás lemezről. Közben lemezhiba ellenőrzése az 1930-as szubrutinnal.
1930—2040:	Lemezhibát ellenőrző szubrutin.
2050—2150:	Várakozó szubrutin.
2160—2230:	Várakozó szubrutin Igen-Nem válaszra.

```
830 REM********************************
840 REM*                              *
850 REM*       SZAMITASOK             *
860 REM*                              *
870 REM********************************
880 PRINT CHR$(147)
890 G=0
900 FOR I=1 TO N:FOR J=1 TO R:T(I)=T(I)+X(I,J)
910 V(J)=V(J)+X(I,J):NEXT J:G=G+T(I):NEXT I
920 REM********************************
930 REM*                              *
940 REM* EREDMENY KIIRASA             *
950 REM*                              *
960 REM********************************
970 PRINT"KEPERNYO(3) VAGY PRINTER(4)";:INPUT ID
980 IF ID<>3 AND ID<>4 THEN 970
990 FOR O=3 TO ID:OPEN O,O:IF O=3 THEN PRINT CHR$(147)
1000 PRINT#O:PRINT#O
1010 PRINT#O," KHI-NEGYZET ANALIZIS"
1020 PRINT#O:PRINT#O
1030 PRINT#O,"          VALAMENNYI REKESZRE (CELLARA)":PRINT#O
1040 PRINT#O," I ESZLELT ERT. ELVART ERT. KHI-NEGYZET"
1050 C2=0:K1=0:FOR J=1 TO R
1060 PRINT#O:PRINT#O,"            ";J;". OSZLOP:":PRINT#O
1070 FOR I=1 TO N:PRINT#O,RIGHT$("    "+STR$(I),3);
1080 K1=K1+1:IF K1<15 OR O=4 OR (J=R AND I=N) THEN 1100
1090 GOSUB 2110:K1=0:PRINT CHR$(147)
1100 E=T(I)*V(J)/G:H=(X(I,J)-E)*(X(I,J)-E)/E:C2=C2+H
1110 EX=INT(E*1E5)/1E5:HX=INT(H*1E5)/1E5
1120 PRINT#O,RIGHT$(SP$+STR$(X(I,J)),12);
1130 PRINT#O,RIGHT$(SP$+STR$(EX),12);
1140 PRINT#O,RIGHT$(SP$+STR$(HX),12)
1150 NEXT I:PRINT#O:NEXT J
1160 PRINT#O,"KHI-NEGYZET      =";C2
1170 PRINT#O
1180 PRINT#O,"SZABADSAGI FOK =";(R-1)*(N-1)
1190 IF O=4 THEN 1220
1200 PRINT#O:PRINT#O
1210 IF O=3 THEN GOSUB 2110
1220 CLOSE O:NEXT O
```

4.3. Normalitásvizsgálat

Normalitásvizsgálat grafikus úton

Elsőként a normalitásvizsgálat egy grafikus módszerét ismertetjük, amely első tájékozódásra mindenképpen ajánlatos, ha a normális eloszlás egyáltalán szóba kerülhet. Végleges döntésre csak abban az esetben alkalmas, ha a múltbeli tapasztalatok szerint normális eloszlást követő változó eloszlását csupán ellenőrizni akarjuk, vagy ha a grafikus eljárás szemmel láthatóan jó eredményre vezet. A vizsgálathoz ún. *Gauss-háló* (Gauss papír) szükséges. Az ilyen koordináta-rendszernek az a tulajdonsága, hogy a normális (Gauss) eloszlás eloszlásfüggvénye ebben a koordinátarendszerben egyenest ad. A vizsgálat elvégzéséhez az eloszlásfüggvényt az empirikus eloszlással közelítjük, azaz mintánkat osztályokba soroljuk és kiszámítjuk az egyes osztályokba eső értékekből a kumulált relatív gyakoriságokat az ismertetett módon.

Az osztályközöket a vízszintes tengelyre (lineáris beosztású tengely) visszük fel, a kumulált relatív gyakoriságokat pedig a függőleges (Gauss beosztású) tengelyre. Ha a pontok jó közelítéssel egy egyenesre esnek, akkor azt mondjuk, hogy a valószínűségi változó eloszlása normális. Azt az abszcissza értéket, amely az 50%-os kumulált relatív gyakorisághoz tartozik (ahol a közelítő egyenes az 50%-os ordinátaértéket metszi), a várható érték becslésének tekinthetjük. A szórás becslését is elvégezhetjük a Gauss-hálón ábrázolt empirikus eloszlásfüggvényből, hiszen a normális eloszlásról tudjuk, hogy a várható érték $\pm 1\sigma$ környezetébe esik a valószínűségi változó értéke 68,3%-os valószínűséggel, $\pm 2\sigma$ környezetébe 95,5%-os valószínűséggel, $\pm 3\sigma$ környezetébe, 99,7%-os valószínűséggel. Annak a valószínűsége, hogy a vizsgált valószínűségi változó értéke a várható értéknél legfeljebb a szórással nagyobb, $50\% + \frac{68,3}{2}\,\% = 84,15\%$. Ezt az ordinátaértéket az egyenesre vetítve, a kapott abszcissza értékből a várható értéket levonva éppen a szórást kapjuk.

Megjegyzendő, hogy ha nem a valódi eloszlásfüggvényről, hanem a minta alapján felvett tapasztalati eloszlásfüggvény ábrájáról olvassuk vissza az 50%-os valószínűségnek megfelelő abszcissza értéket, az szigorúan véve a medián értéket adja.

Nézzünk most egy példát a Gauss-háló használatára. Tegyük fel, hogy van n mérési adatunk, és feltevésünk szerint normális eloszlásúak a mérési adatok. Rendezzük nagyság szerinti sorrendbe mérési adatainkat:

$$x_1 \leqq x_2 \leqq \ldots \leqq x_n.$$

Minden egyes mérési pontra számítsuk ki a kumulált relatív gyakoriságot (a tapasztalati eloszlásfüggvény x_i-beli értékét)

$$F(x_j) = \frac{\sum\limits_{x_i \leqq x_j} 1}{n+1} = \frac{j}{n+1}.$$

(A nevezőbe azért írtunk n helyett $n+1$-et, hogy az utolsó mérési pontot is ábrázolni tudjuk a Gauss-hálón.) Válasszunk megfelelő léptéket x-nek, és ábrázoljuk ennek függvényében a kiszámított valószínűségeket.

Tegyük fel, hogy B 100-as betonból készült próbatestek nyomószilárdságát N-ban mérjük, és 30 véletlenszerűen kiválasztott próbahengerre a következő adatokat kapjuk:

999,1	1003,2	1002,1	999,2	989,7	1006,7
1012,3	996,4	1000,2	995,3	1008,7	993,4
998,1	997,9	1003,1	1002,6	1001,8	996,5
992,8	1006,5	1004,5	1000,3	1014,5	998,6
989,4	1002,9	999,3	994,7	1007,6	1000,9

Az adatokat sorbarendezve és a kumulált relatív gyakoriságokat kiszámítva a következő táblázathoz jutunk:

i	x_i	$F(x_i)$	i	x_i	$F(x_i)$	i	x_i	$F(x_i)$
1	989,4	0,032	11	998,6	0,355	21	1002,9	0,678
2	989,7	0,064	12	999,1	0,386	22	1003,1	0,710
3	992,8	0,097	13	999,2	0,418	23	1003,2	0,742
4	993,4	0,129	14	999,3	0,450	24	1004,5	0,774
5	994,7	0,161	15	1000,2	0,483	25	1006,5	0,805
6	995,3	0,193	16	1000,3	0,516	26	1006,7	0,870
7	996,4	0,225	17	1000,9	0,549	27	1007,6	0,902
8	996,5	0,257	18	1001,8	0,581	28	1008,7	0,908
9	997,9	0,290	19	1002,1	0,613	29	1012,3	0,934
10	998,1	0,322	20	1002,6	0,645	30	1014,5	0,977

A táblázat adatait a 17. ábrán látható Gauss-hálón ábrázoltuk. A pontokra szemmel láthatóan egyenest lehet illeszteni. Ez azt jelzi, hogy mérési adataink normális eloszlásúak. Másrészt, az egyenest „szemre" behúzva, az 50%-os kumulált relatív gyakoriságnál leolvasott x érték 1000,6, ami a mérési adatok várható értékének tekinthető, és a 84,15%-os és 50%-os függvényértékek különbségéből visszaolvasva a szórás 6,2-nek adódik.

Normalitásvizsgálat χ^2 próbával

Ha a Gauss-hálóval végzett normalitásvizsgálat nem ad egyértelmű eredményt, és a minta elég nagy (minden osztályközbe 5-nél több adatunk esett), akkor a normalitásvizsgálatot χ^2 próbával is el lehet végezni. (A χ^2 próba sok más statisztikai vizsgálatra is jó segédeszköz — illeszkedésvizsgálat, homogenitás- és függetlenségvizsgálat végezhető vele, l. az előző pontot.) A χ^2 próbával végzett normalitásvizsgálat az ún. illeszkedésvizsgálatok közé tartozik. Ilyenkor azt vizsgáljuk, hogy valamely valószínűségi változó eloszlása tekinthető-e egy adott konkrét — vagy egy adott típusú — eloszlásnak. Ha a nullhipotézis az eloszlásfüggvényt egyértelműen meghatározza, akkor ez az ún. *tiszta illeszkedésvizsgálat.* Amikor a hipotézis az eloszlásfüggvényeknek csupán a típusára vonatkozik, és az eloszlás paramétereit a mintából kell becsülni, akkor *becsléses illeszkedésvizsgálatról* beszélünk.

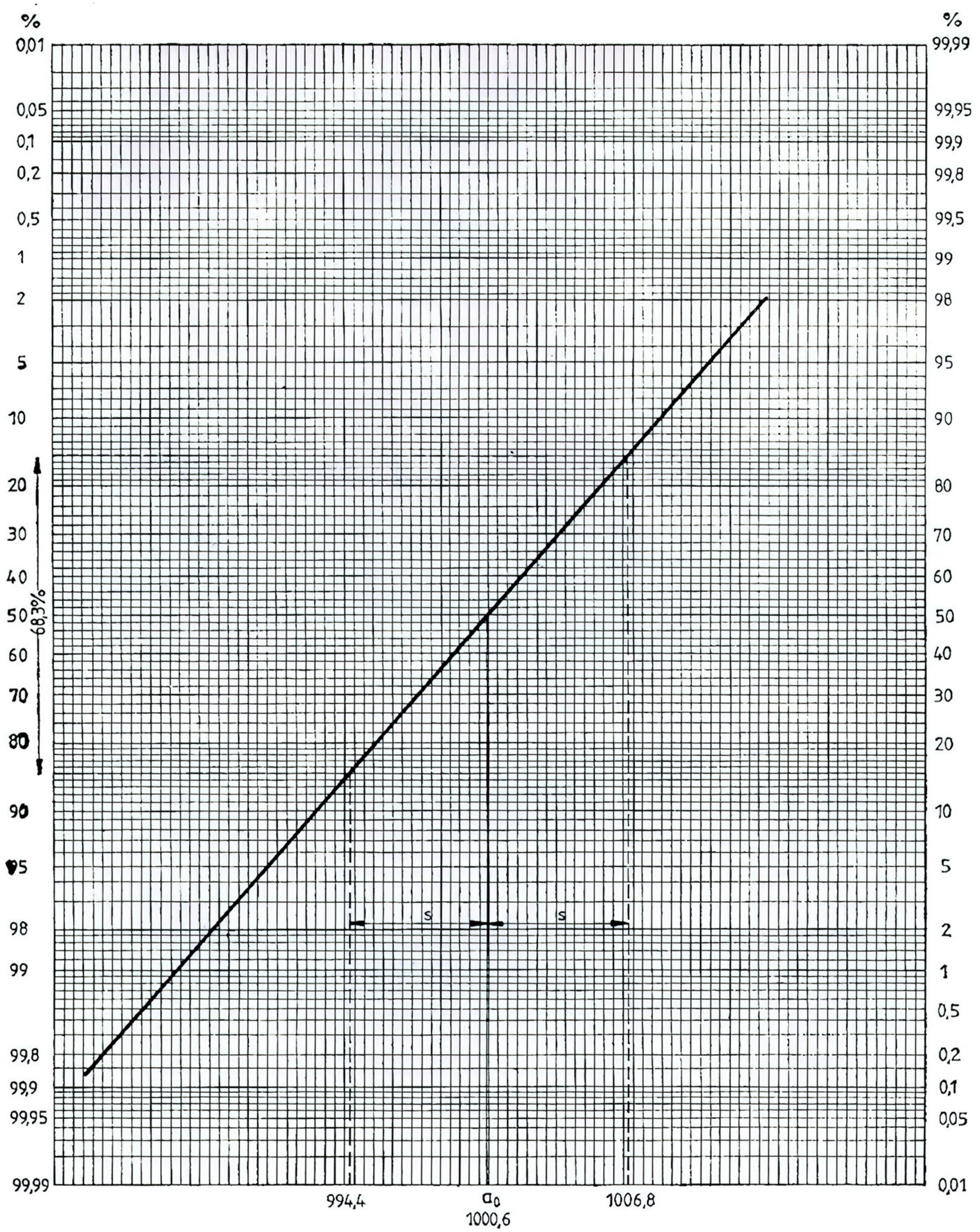
A χ^2 próbával illeszkedésvizsgálat mind diszkrét, mind folytonos eloszlású valószínűségi változó esetén végezhető. Mi itt csak normalitásvizsgálattal, így a folytonos eloszlású valószínűségi változó esetére vonatkozó illeszkedésvizsgálattal foglalkozunk.

Soroljuk adatainkat m osztályba. Tekintsük azokat a diszkrét eseményeket, hogy a valószínűségi változó értéke az r-edik osztályközbe esik.

$$A_r = \{z_{r-1} \leqq \xi < z_r \quad r = 1, 2, \ldots, m\}.$$

Vizsgáljuk meg azt a hipotézist, hogy

$$H_0 \colon P(A_r) = p_r = F(z_r) - F(z_{r-1})$$



17. ábra. Normalitásvizsgálat Gauss-háló segítségével

azaz azt, hogy az A_r esemény valószínűsége az ismertnek feltételezett folytonos eloszlás eloszlásfüggvényének z_r és z_{r-1} argumentumához tartozó értékeinek különbsége.

Konstruáljuk meg a következő statisztikát:

$$\chi^2 = \sum_{r=1}^{m} \frac{(f_r - Np_r)^2}{Np_r},$$

ahol f_r az egyes osztályközökhöz tartozó gyakoriságok értéke, p_r az eloszlásfüggvény táblázatából leolvasott két osztályközhatár valószínűségének különbsége, $N = \sum_{r=1}^{m} f_r$ a minta elemszáma.

Ha a nullhipotézis teljesül, akkor a számlálóban a zárójeleken belül a gyakoriságok várható értéktől való eltérése áll. Tehát minden zárójelben 0 várható értékű valószínűségi változó áll, e körül mutatva véletlen ingadozásokat. Be lehet bizonyítani, hogy ha $N \to \infty$, akkor ez a kifejezés az $(m-1)$ paraméterű χ^2 eloszlás felé tart. Ily módon, ha m rögzített és N elég nagy, a kifejezés közel χ^2_{m-1} eloszlású, s így a nullhipotézis vizsgálatára adott $(1-\varepsilon)$ szinthez a következő kritikus tartományt adhatjuk:

$$X_k = \{\chi^2 \geqq \chi^2_{m-1}\}.$$

A kritikus értékeket az $\varepsilon = 0{,}01$, ill. $\varepsilon = 0{,}05$ (1, ill. 5%-os) szignifikanciaszinthez a Melléklet táblázata tartalmazza különböző paraméterek $(m-1)$ esetén.

Ha a próbát a normális eloszlás ellenőrzésére használjuk, akkor a p_r valószínűségek kiszámításához a normális eloszlásfüggvényt használjuk (l. a Melléklet 1. táblázatát). Ha az eloszlás paramétereit (normális eloszlás esetén szórását és várható értékét) nem ismerjük, s így a mintából becsüljük, akkor a χ^2 statisztika eloszlása nem $(m-1)$, hanem a mintából becsült paraméterek számával kevesebb (normális eloszlás becsléses illeszkedésvizsgálata esetén $(m-3)$) paraméterű.

A módszer illusztrálására tekintsük ismét az előző pontban ismertetett példát. A 30 mérési adat „szemre", a Gauss-háló alkalmazásával normális eloszlásúnak látszott. Kérdés: igazolható-e ez az állítás χ^2 próbával is?

Válasszunk 5 N szélességű osztályokat. A felső osztályhatárok és az egyes osztályokhoz tartozó gyakoriság értékek így a következők lesznek:

OSZTÁLYKÖZ	GYAKORISÁG
— 995	5
995,1—1000	9
1000,1—1005	10
1005,1—1010	4
1010,1—1015	2
$\bar{x} = 1000{,}6$	30
$s = 5{,}976$	

Az egyszerűség kedvéért fogadjuk el, hogy az adatok átlaga 1000,6, szórása 6,2 N.

A példát megoldhatjuk a STATISZTIKAI PRÓBÁK menüjéből a NORMALITÁSVIZSGÁLAT KHI-NÉGYZET PRÓBÁVAL (5) program kiválasztásával. A program lehetőséget ad (az ismertetés felkínálása után) lemezről vagy billentyűzetről történő adatbevitelre. Akár a teljes adatsor (nyers adatok), akár a már osztályba sorolt adatok bevitelére. Válasszuk most az utóbbit, és vigyük be az előző táblázat szerint osztályba sorolt nyomásszilárdság adatainkat. A minta elemszáma 30, átlag-

értéke 1000,6, a korrigált tapasztalati szórást 6,2 értékűnek tekintjük. A kérdésekre értelemszerűen válaszolva beütjük az egyes osztályközök felső határait és az illető osztály gyakoriságát. A képernyőt kijelölve output perifériának, a következő eredmény jelenik meg:

KHI-NÉGYZET ÉRTÉKE	0,4015
SZABADSÁGI FOK	2
A VALÓSZÍNŰSÉG	0,7819

Utalva a χ^2 próbával kapcsolatban korábban elmondottakra, az eredményt úgy értelmezhetjük, hogy az adatok normális eloszlásúak, hiszen az erre vonatkozó nullhipotézisünket csak 22%-os szignifikanciaszint felett tudnánk elvetni. A szokásos 1 és 5%-os szignifikanciaszinten a minta normális eloszlásúnak mondható.

NORMALITÁSVIZSGÁLAT KHI-NÉGYZET PRÓBÁVAL

A változók listája:

N	— a mintacsoport elemszáma,
X(I)	— a mintacsoport elemei,
S1	— az elemek összegzése,
M1	— az elemek átlaga,
R1	— részeredmény,
Q1	— korrigált tapasztalati szórás,
D	— az osztályközök száma,
L(1)	— az osztályközök felső határai,
K(I)	— a gyakoriságok összege a j-edik osztályközig,
P(I), R(I)	— a gyakoriságok,
F(J), G(J)	— részeredmények,
E, Z	— normális eloszlás számításánál a változók,
Q(J)	— normális eloszlásnál az eloszlásfüggvény értéke,
W, WW	— khi-négyzet eloszlásnál a szabadsági fok,
J, A, B, S, T	— khi-négyzet eloszlás számításánál a változók,
OX	— adatmátrixból kijelölt oszlopok száma,
A$	— Igen-Nem választás,
ID	— kiíratásnál választás (képernyő vagy nyomtató),
N$	— adatfile neve,
HS, HN$, S, SZ	— lemezhiba ellenőrzése.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—250:	A program neve.
260—1780:	Adatbevitel, számítások.
1790—2020:	Eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
2030:	Visszatérés a FŐMENÜ c. programra.
2040—2240:	A program által nyújtott lehetőségek rövid ismertetése.
2250—2600:	Adatbeolvasás lemezről. Közben lemezhiba ellenőrzése a 2610-es szubrutinnal.
2610—2720:	Lemezhibát ellenőrző szubrutin.
2730—2830:	Várakozó szubrutin.
2840—2910:	Várakozó szubrutin Igen-Nem válaszra.

```
570 REM******************************
580 REM*                            *
590 REM* AZ ADATOK SORBARENDEZESE   *
600 REM*                            *
610 REM******************************
620 PRINT CHR$(147):PRINT"SORBARENDEZES"
630 FOR I=1 TO N-1:CS=0:PRINT"[illegible]";I
640 FOR J=1 TO N-I:PRINT"[illegible]";J
650 IF X(J)>X(J+1) THEN W=X(J):X(J)=X(J+1):X(J+1)=W:CS=1
660 NEXT J:IF CS=0 THEN I=N
670 NEXT I
680 REM******************************
690 REM*                            *
700 REM* OSZTALYKOZOKBE SOROLAS     *
710 REM*                            *
720 REM******************************
730 PRINT:PRINT"A MINTACSOPORT LEGKISEBB ELEME :":PRINT X(1)
740 PRINT:PRINT"A MINTACSOPORT LEGNAGYOBB ELEME :":PRINT X(N)
741 IF X(1)<>X(N) THEN 750
742 PRINT:PRINT"MINIMUM=MAXIMUM.AZ OSZTALYBA SOROLASNAK SEMMI ERTELME."
743 GOSUB 2790:GOTO 260
750 PRINT:PRINT"HANY OSZTALYKOZBE KIVANJA SOROLNI AZ     ADATOKAT?"
760 INPUT D
770 IF D>10 OR D<2 OR D<>INT(D) THEN PRINT"2 ES 10 KOZTI EGESZ!":GOTO 760
780 L(0)=X(1):L(D)=X(N)
790 HS=0:FOR I=1 TO D-1
800 PRINT"ADJA MEG A";I;".OSZTALYKOZ FELSO HATARAT!":INPUT L(I)
810 IF L(I)<=L(I-1) OR L(I)>L=(D) THEN PRINT"TEVES ADAT!":HS=1:I=D:GOTO 850
820 IF I<3 THEN 850
830 IF ABS((L(I)-L(I-1))/(L(I-1)-L(I-2))-1)<1E-3 THEN 850
840 PRINT"EGYENLO OSZTALYKOZOKET ADJON MEG!":I=D:HS=1
850 NEXT I:IF HS<>0 THEN 790
860 IF L$="I" OR C$="I" THEN 930
870 PRINT:K=0:FOR I=1 TO D:PRINT"ADJA MEG A";I;". OSZTALYKOZ GYAKORISAGAT!"
880 INPUT P
890 IF P<>INT(P) OR P<0 OR P>N THEN PRINT"0 ES";N;" KOZTI EGESZ!":GOTO 880
900 P(I)=P:K=K+P:K(I)=K:NEXT I
910 IF K<>N THEN PRINT"A GYAKORISAGOK OSSZEGE NEM";N;"!":GOTO 860
920 GOTO 970
930 K(0)=0:K(D)=N
940 FOR J=1 TO D-1:FOR I=K(J-1)+1 TO N
950 IF X(I)>L(J) THEN K(J)=I-1:I=N
960 NEXT I:NEXT J
970 H=0:FOR I=1 TO D:P(I)=K(I)-K(I-1):IF P(I)<8 THEN H=1
980 NEXT I
990 IF H=0 THEN 1070
1000 PRINT"VAN OLYAN OSZTALYKOZ, AHOL 8-NAL          KEVESEBB ELEM VAN."
1010 PRINT"KIVAN JAVITANI";:GOSUB 2890:IF A$="I" THEN 750
1020 REM******************************
1030 REM*                            *
1040 REM* AZ ADATOK KIIRASA          *
1050 REM*                            *
1060 REM******************************
1070 PRINT:PRINT"KIIRJUK AZ ADATOKAT";:GOSUB 2890:PRINT
1080 IF A$="N" THEN 1380
1090 PRINT"KEPERNYO(3) VAGY PRINTER(4) ";:INPUT ID
1100 IF ID<>3 AND ID<>4 THEN 1090
1110 FOR O=3 TO ID:OPENO,O:IFO=3 THEN PRINT CHR$(147)
1120 PRINT#O
1130 PRINT#O,"NORMALITASVIZSGALAT "
1140 PRINT#O,"  KHI-NEGYZET PROBAVAL"
1150 PRINT#O
1160 IF L$="N" AND C$="N" THEN 1280
1170 PRINT#O:PRINT#O,"A RENDEZETT ADATSOR:"
1180 SN=3:I1=1:K1=6:IF O=4 THEN SN=6
1190 I2=I1+SN-1:IF N<I2 THEN I2=N
1200 K1=K1+1:IF K1<15 OR O=4 OR I2=N THEN 1240
```

```
1210 PRINT
1220 PRINT"KERI A TOBBI ADATOT IS";:GOSUB 2890:IF A$="N" THEN I2=N:GOTO 1250
1230 K1=0:PRINT CHR$(147)
1240 FOR J=I1 TO I2:PRINT#O,RIGHT$(SP$+STR$(X(J)),12);:NEXT J:PRINT#O
1250 IF I2<>N THEN I1=I1+SN:GOTO 1190
1260 PRINT#O:IF O=3 THEN GOSUB 2790
1270 IF O=3 THEN PRINTCHR$(147)
1280 PRINT#O:PRINT#O,"OSZTALYKOZ                     GYAKORISAG"
1290 FOR I=1 TO D:X$=SP$:IF I>1 THEN X$=RIGHT$(SP$+STR$(L(I-1)),12)
1300 Y$=SP$:IF I<D THEN Y$=RIGHT$(SP$+STR$(L(I)),12)
1310 PRINT#O,X$;" - ";Y$;RIGHT$(SP$+STR$(K(I)-K(I-1)),8):NEXT I
1320 PRINT#O:IF O=3 THEN GOSUB 2790
1330 CLOSE O:NEXT O
1340 REM*****************************
1350 REM*                           *
1360 REM* SZAMITASOK
1370 REM*                           *
1380 REM*****************************
1390 PRINT CHR$(147):PRINT"SZAMITASOK"
1400 IF Q1>=0 THEN 1430
1410 M1=0:FOR I=1 TO N:M1=M1+X(I):NEXT I:M1=M1/N
1420 Q1=0:FOR I=1 TO N:Q1=Q1+(X(I)-M1)*(X(I)-M1):NEXT I:Q1=SQR(Q1/(N-1))
1430 FOR J=1 TO D-1
1440 FJ=1E37:IF Q1<>0 THEN F(J)=(L(J)-M1)/Q1
1450 NEXT J
1460 FOR J=1 TO D-1
1470 E=EXP(-F(J)↑2/2)/2.5066283
1480 Z=F(J)
1490 F(J)=1/(1+.33267*ABS(F(J)))
1500 Q(J)=1-E*(.4361836*F(J)-.1201676*F(J)↑2+.937298*F(J)↑3)
1510 IF Z>=0 THEN 1530
1520 Q(J)=1-Q(J)
1530 NEXT J
1540 Q(0)=0:Q(D)=1
1550 FOR J=1 TO D
1560 G(J)=Q(J)-Q(J-1)
1570 NEXT J
1580 C=0
1590 FOR J=1 TO D
1600 C=C+(P(J)-N*G(J))↑2/N/G(J)
1610 NEXT J
1620 W=D-3
1630 J=1
1640 FOR I=W TO 2 STEP -2
1650 J=J*I
1660 NEXT I
1670 A=C↑(INT((W+1)/2))*EXP(-C/2)/J
1680 IF INT(W/2)=W/2 THEN 1710
1690 B=SQR(2/C/3.1415927)
1700 GOTO 1720
1710 B=1
1720 S=1:T=1
1730 WW=W
1740 W=W+2
1750 T=T*C/W
1760 IF T<.00001 THEN 1840
1770 S=S+T
1780 GOTO 1740
```

4.4. A várható érték vizsgálata az u és t statisztika használatával

Ebben a fejezetben mindig feltételezzük, hogy normális eloszlású alapsokaságból származó véletlen minta (vagy minták) alapján arról kell döntenünk, hogy az ismeretlen várható érték milyen feltételezett értékkel egyenlő adott (előre rögzített) valószínűséggel.

A sokaság normális eloszlásának feltételezését az előző fejezetben szereplő normalitásvizsgálatok egyikével lehet ellenőrizni. A normális eloszlás feltételezése azonban ne riassza el a felhasználót a fejezetben szereplő próbák alkalmazásától. Van egy alapvető statisztikai tétel, az ún. központi határeloszlás tétele. Ez kimondja, hogy ha n elemű véletlen mintákat veszünk egy olyan eloszlású sokaságból, amelynek M a várható értéke és σ a szórása (eloszlásáról nem tételezve fel semmit), akkor a minták számtani átlagának, $\bar{x}$-nak az eloszlása közelítően normális eloszlású lesz, ugyanazzal az M várható értékkel és $\sigma/\sqrt{n}$ szórással. A normális eloszláshoz való közelítés a minták elemszámának növelésével javul.

Ebből a gondolatsorból azt kell kiemelnünk, hogy elegendően nagy elemszámú mintánál (pl. $n \geqq 20$) a minták számtani átlagának eloszlása közelítően normális eloszlású, akármilyen is volt az alapsokaság eredeti eloszlása.

Mivel a mintaátlagok várható értéke az eredeti sokaság várható értéke, a várható értéket a közel normális eloszlású mintaátlagok várható értékének lehet tekinteni. Az $(\bar{x}-M)$ valószínűségi változó 0 várható értékű, szórása $\sigma/\sqrt{n}$. Eloszlása normális, ha az eredeti sokaságé normális volt, és közel normális, akármilyen is volt az eredeti sokaság eloszlása feltéve, hogy a minta véletlenszerű és elegendően nagy elemszámú. Így a most bemutatandó statisztikai próbák egzaktnak mondhatók, ha a sokaság eloszlása normális volt, és közelítő jelleggel akkor is alkalmasak statisztikai vizsgálatokra, ha a sokaság eloszlásáról semmit sem tudunk.

Ebben a fejezetben a várható értékek egyezésének vizsgálatára az u-próbát és a t-próbát (Student próba) mutatjuk be és szót ejtünk a Student próba egy módosult változatának tekinthető Welch próbáról is.

Mind az u-, mind a t-próba lehet egymintás és kétmintás, azaz alkalmas a várható értékre vonatkozó feltételezés ellenőrzésére, ill. két valószínűségi változó várható értékei egyezésének elbírálására. A hipotézisvizsgálat elvégzése függ a nullhipotézisen kívül az ellenhipotézistől is. Mit is értünk ez alatt? Általánosságban a következőket mondhatjuk: A nullhipotézis legtöbbször egy ismeretlen paraméter és egy számérték egyenlőségére vonatkozik. Pl.:

$$H_0:\ M(\alpha) = a.$$

Ha az ellenhipotézis csupán a nullhipotézis tagadása, azaz

$$H_1:\ M(\alpha) \neq a$$

az azt jelenti, hogy mindegy számunkra, hogy az ismeretlen várható érték nagyobb-e, vagy kisebb, mint a. Ezért az előírt p szignifikanciaszintet úgy kell értelmeznünk, hogy az illető eloszlás sűrűségfüggvényének mindkét végéből $p/2$-nek megfelelő területet elhagyunk. Az így kapott határokat jelölő kritikus értékkel kell összehason-

lítani a számított próbastatisztikát. Ez az ún. kétoldali próba, ill. kétoldali (általában szimmetrikus) intervallum kijelölése.

Ha az ellenhipotézis

$$M(\varepsilon) > a$$

alakú, akkor a nullhipotézist csak abban az esetben kell elvetnünk, ha a próbastatisztika a kitűzött (kritikus) értéknél nagyobb pozitív szám. Ezért az eloszlás sűrűségfüggvényének csak az egyik oldalán kell kritikus tartományt kijelölni, azaz a p szignifikanciaszint az egyik oldalon levágott terület mértéke. Ez az ún. egyoldali próba.

Ha az ellenhipotézis

$$M(\varepsilon) < a,$$

akkor az előbb elmondottak értelemszerűen megfordítva (előjelcserével) továbbra is érvényesek maradnak. A megbízhatósági szint minden esetben $(1-p)$, vagy százalékban kifejezve $100(1-p)\%$.

4.4.1. Egymintás u-próba

Ezt a próbát akkor alkalmazzuk, ha egy ismert σ elméleti szórású normális eloszlású alapsokaságból vett n elemű minta átlaga alapján a valószínűségi változó várható értékére vonatkozó nullhipotézisünket akarjuk ellenőrizni. Tudjuk, hogy a normális eloszlás két paramétere az M várható értéke és a σ szórása. Mivel σ-t ismertnek tételezzük fel, az M várható értékről feltételezzük, hogy egyenlő a_0-lal, azaz azt akarjuk megtudni, helyes-e az a nullhipotézis, hogy a várható érték a_0.

$$H_0:\ M = a_0$$

vagy másképpen felírva:

$$H_0:\ a_0 - M = 0.$$

Ha H_0 igaz, akkor az

$$u = \frac{\bar{x} - a_0}{\dfrac{\sigma}{\sqrt{n}}}$$

statisztika, mint a statisztikai mintából képzett valószínűségi változó normális eloszlású. 0 várható értékkel és 1 szórással. A nullhipotézis ellenőrzésére a standard normális eloszlás táblázatából választunk egy u_p számot, amely egy adott p megbízhatósági szinthez (szignifikanciaszinthez) tartozik. Ha a szignifikanciaszintet a szokásos 0,05 (5%)-ban rögzítjük, az azt jelenti, hogy ennyi kockázatot vállalunk arra, hogy egy jó hipotézist elvetünk.

Ha az ellenhipotézis a

$$H':\ M \neq a_0,$$

akkor kétoldali próbáról van szó, ezért a 0,05-ös szignifikanciaszintet is ennek megfelelően kell értelmezni, azaz képzeletben az eloszlás sűrűségfüggvényéből mindkét oldalon 2,5—2,5%-ot kell levágni. Így a kritikus értéket a 0,025-ös szignifikanciaszinthez kell a Melléklet 1. táblázatából kikeresni. $p=0{,}025$-höz 1,96-os kritikus érték tartozik.

Ha a számított u érték ennél nagyobb pozitív vagy az ellentettjénél kisebb negatív érték, a hipotézist elvetjük. Azaz, ha

$$|u| > u_{\text{krit}},$$

akkor a hipotézist (H_0) hamisnak tekintjük.

Például egy folyóvízben a higanykoncentrációt sokszor megmérve 2,0 μg/l-nek találták és sok (elvileg végtelen sok) mérés alapján tudjuk, hogy $\sigma = 0{,}04$ μg/l. Adott időben 64 mintát véve, azokat elemezve, a higanykoncentrációt 1,985 μg/l-nek találjuk. Ellenőrizzük azt a hipotézist, hogy a folyóvíz higanytartalmának várható értéke 2,0 μg/l. Oldjuk meg a feladatot az EGYMINTÁS U-PRÓBA program használatával!

A főmenüből a STATISZTIKAI PRÓBÁK ágat kell választani, a STATISZTIKAI PRÓBÁK menüjéből az U-STATISZTIKA HASZNÁLATÁ-t.

Ennek hatására megjelenik az U-STATISZTIKA menüje:
EGYMINTÁS U-PRÓBA (E)
KÉTMINTÁS U-PRÓBA (K)
ISMERTETÉST KÉREK (I)
VÉGE (V)

Ebből természetesen az egymintás u-próbát választjuk. Az adatokat billentyűzetről is bevihetjük vagy adatfile-ból olvashatjuk be. Mivel nem a primer adatokkal dolgozunk, hanem ismerjük az átlagot és a szórást, egyszerűbb billentyűzetről bevinni az adatokat. Ezt választva interaktív üzemmódban megkérdezi a program a minta elemszámát (64), az átlagot (1,985) és a szórást (0,04). Ezek megadása után megjelenik a következő menü a képernyőn.

AZ ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:
U-STATISZTIKA ÉRTÉKEINEK SZÁMÍTÁSA (U)
MEGBÍZHATÓSÁGI INTERVALLUM KISZÁMÍTÁSA (M)
A MÉRÉSEK SZÁMÁNAK MEGHATÁROZÁSA (N)
VISSZATÉRÉS A FŐMENÜHÖZ (F)
VÉGE (V)

Először válasszuk az u-érték kiszámítása programágat. Ekkor a program megkérdezi a várható értéket (2,0), és azt, hogy egyoldali vagy kétoldali próbáról van-e szó. Mivel most ellenhipotézisünk

$$H'\colon\ M(\bar{x}) \neq a_0,$$

a kétoldali próbát kell választanunk.

Ezt követően megkérdezi a program a megbízhatósági szintet. Legyen a szignifikanciaszint a szokásos 5%, azaz a megbízhatósági szint 95%. A kritikus érték 5%-os szignifikanciaszinten a Melléklet 1. táblázata szerint 1,96. Ennek megadása után az eredményt képernyőre kiíratva a következő jelenik meg

$$\text{U-ÉRTÉK} = -2{,}999.$$

95%-os megbízhatósági szinten a hipotézist el kell vetnünk, hiszen a statisztika számított értéke nagyobb, mint a kritikus érték.

Tetszőleges billentyű megnyomásával újra visszatérhetünk az egymintás u-próba menüjéhez. Előbb azonban vizsgáljuk meg, mi történne:

a) Ha a szórás az alapsokaságban nem 0,04, hanem 0,08 lett volna?

b) Ha a kétoldali próba helyett egyoldali próbát alkalmaznánk, azaz a nullhipotézist a

$$H': M(\bar{x}) < a_0$$

ellenhipotézissel szemben vizsgálnánk?
c) Ha a szignifikanciaszintet (a kétoldali próbánál) 0,02-re, azaz 2%-ra csökkentenénk?

A válaszok megkeresését a programmal az Olvasóra bízzuk. Támpontul a helyes válaszok.
a) Ha a szórás az alapsokaságban nem 0,04, hanem 0,08 lett volna, a nullhipotézist el kellene fogadnunk.
b) A nullhipotézist egyoldali próbánál is el kellene vetnünk.
c) Ha a szignifikanciaszintet csökkentjük, azaz kisebb kockázatot vállalunk az igaz hipotézis elvetésére, akkor a hipotézist esetleg elfogadhatjuk. Ez esetünkben 0,002-es, azaz 0,2%-os szignifikanciaszintnél teljesül. 2%-os szignifikanciaszinten azonban még el kell utasítanunk a nullhipotézist.

4.4.2. Kétmintás u-próba

Ha két normális eloszlású, ismert szórású sokaságokból vett minták alapján a két sokaság várható értékének egyezésére vonatkozó hipotézisünket akarjuk ellenőrizni, kétmintás u-próbát alkalmazunk.

Legyen ξ és η a két valószínűségi változónk, $\bar{x}$ és $\bar{y}$ a ξ-re és η-ra vett minták aritmetikai közepe, $\sigma_1^2=\sigma_2^2=\sigma^2$ a valódi (ismert) szórásnégyzet. A

$$H_\emptyset: M(\xi) = M(\eta)$$

nullhipotézist a ξ-re vett n elemű és az η-ra vett m elemű minták alapján kívánjuk ellenőrizni. A próbastatisztika:

$$u = \frac{\bar{x}-\bar{Y}}{\left(\frac{\sigma^2}{n}+\frac{\sigma^2}{m}\right)^{1/2}}.$$

Ha $H_\emptyset$ igaz, akkor u eloszlása $N(0,1)$ és adott ε-hoz a

$$H_1: M(\xi) \neq M(\eta)$$

ellenhipotézis mellett a kritikus tartomány:

$$X_k = \{u \leqq -u_\varepsilon \text{ vagy } u \geqq u_\varepsilon\}$$

ε szignifikanciaszinten (kétoldali próba). A H_1': $M(\xi) > M(\eta)$ ellenhipotézis esetén pedig az

$$X_k': \{u \leqq u_\varepsilon\}$$

a kritikus tartomány.

A kétmintás u-próba alkalmazásaként tekintsük a következő egyszerű példát: Legyen A és B két abszorpciós torony, amelyeket azonos nitrózusgáz elnyeletésére alkalmaznak, azonos nitrogén-oxid-koncentrációjú hulladékgázok nitrogén-oxidmentesítésére. Az A oszlop kimenő gáza 9 független mérés szerint átlagosan 0,1672 g/m³, a B oszlopé 16 független mérés szerint 0,1683 g/m³ nitrogén(II)-oxidot

tartalmaz. A kimenő gázok koncentrációja stacionárius üzemmenet esetén állandó, a nitrogén-oxid-mérés szórása $\sigma=0{,}00129$ g/m³. Kérdés, hogy $1-\varepsilon=95\%$-os megbízhatósági szinten azonosnak fogadható-e el a két elnyelető torony kilépő gázának nitrózusgáz-tartalma? A feladat az $M(\xi)=M(\eta)$ nullhipotézis vizsgálata a

$$H': M(\xi) \neq M(\eta) \quad \text{ellenhipotézissel szemben.}$$

Mivel az u-próbához szükséges feltételek fennállnak, a kétmintás u-próbát alkalmazzuk.

Hívjuk be a STATISZTIKAI PRÓBÁK menüjéből az U-STATISZTIKA ALKALMAZÁSA programot, annak menüjéből kiválasztjuk a KÉTMINTÁS U-PRÓBÁ-t, és a képernyőt input perifériának kijelölve a képernyőn megjelenő kérdésekre válaszolva rendre megadjuk a két minta elemszámát, átlagértékét és szórását, és megadjuk, hogy kétoldali próbáról van szó, 95%-os megbízhatósági szinten. A képernyőn megjelenő eredmény a következő lesz:

U-ÉRTÉK = −2,046.
KÉTOLDALI PRÓBA ESETÉN 95%-os MEGBÍZHATÓSÁGI SZINTEN A HIPOTÉZIST EL KELL VETNI.

Valóban, az u aktuális értéke a $\{-1{,}96; 1{,}96\}$ tartományon kívül, tehát a kritikus tartományba esik, ezért logikus, hogy a nullhipotézist, vagyis a két elnyeletés hatékonyságának azonosságát elvetjük. Kérdés azonban, hogy milyen szignifikanciaszinten kellene a nullhipotézist elvetnünk? Ennek vizsgálatára hívjuk be az ELOSZLÁSOK menüjéből a NORMÁLIS ELOSZLÁS programágat. (Tudjuk, hogy az u-statisztika standard normális eloszlású, ha H_0 nullhipotézis igaz.)

A változó értékéül adjuk meg az imént kiszámított −2,046 értéket. A képernyőn megjeleníthetjük az eredményt:

A SŰRŰSÉGFÜGGVÉNY ÉRTÉKE	0,04914,
AZ ELOSZLÁSFÜGGVÉNY ÉRTÉKE	0,02036.

A 0,02036 eloszlásfüggvény-érték úgy értendő, hogy ennyi a valószínűsége a −2,0465-nél kisebb értékeknek standard normális eloszlású valószínűségi változó esetén. Természetesen ugyanennyi a valószínűsége a 2,0465-nél nagyobb értékeknek is, összesen 0,04072 a valószínűsége annak, hogy valószínűségi változó a $\{-2{,}0465, +2{,}0465\}$ tartományon kívül esik.

Következésképpen $100\,(1-0{,}04072)=95{,}928\%$-os, azaz durván 96%-os megbízhatósági szinten már különbséget lehetne tenni a két elnyelető torony hatékonysága között! Győződjünk meg állításunk helyességéről! Végezzünk most ismét kétmintás u-próbát, de ez alkalommal 96%-os megbízhatósági szinten! A próba elvégzését az Olvasóra bízzuk.

U-PRÓBA

A változók listája:

N, M	— mintacsoportok elemszáma,
X(I), Y(I)	— mintacsoportok elemei,
S1, S2	— elemek összegzése,
M1, M2	— elemek átlaga,
Q1, Q2	— szórások,
P	— várható érték,

U — u-érték,
F — konfidencia intervallum (számított),
R — konfidencia intervallum (megadott),
G — mérések száma,
D — megbízhatósági szint,
L — segédváltozó,
K(L) — u kritikus értékei,
W — egy- vagy kétoldali hipotézisnél választás,
OX, OY — adatmátrixból kijelölt oszlopok száma,
D$ — menünél választás,
H$ — Igen-Nem választás,
B$ — egymintás u-Próba menüjénél választás,
ID — kiíratásnál választás (képernyő vagy nyomtató),
N$ — adatfile neve,
HS, HN$, S, SZ — lemezhiba ellenőrzése.

Programmagyarázat:

100: Programcsomag-szervezés.
110—240: A program neve.
250—270: DATA: u kritikus értékei.
280: U kritikus értékeinek beolvasása.
290—420: Menü kiírása, választás, elágazások.
430—1290: Adatbevitel.
1300—1530: Egymintás u-próba menüjének kiírása, választás, elágazások.
1540—1560: U-érték kiszámítása.
1570—1590: Megbízhatósági intervallum kiszámítása.
1600—1650: Mérések számának meghatározása.
1660: Kétmintás u-próba számítása.
1670—1720: Egyoldali vagy kétoldali hipotézisre választás, elágazás.
1730—1790: Szubrutin megbízhatósági szint választására, egyoldali hipotézis esetén.
1800—1850: Szubrutin megbízhatósági szint választására kétoldali hipotézis esetén.
1860—2330: Eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
2340—2580: A program által nyújtott lehetőségek rövid ismertetése.
2590: Visszatérés a FŐMENÜ c. programra.
2600—3050: Adatbeolvasás lemezről és a vizsgálandó mintacsoportok kijelölése. Közben lemezhiba ellenőrzése a 3060-as szubrutinnal.
3060—3170: Lemezhibát ellenőrző szubrutin.
3180—3280: Várakozó szubrutin.
3290—3360: Várakozó szubrutin Igen-Nem válaszra.

```
1540 INPUT"ADJA MEG A VARHATO ERTEKET!  ";P
1550 U=1E37*SGN(M1-P):IF Q1<>0 THEN U=(M1-P)/(Q1/SQR(N))
1560 GOTO 1670
1570 GOSUB 1800
1580 F=K(L)*Q1/SQR(N)
1590 GOTO 1860
1600 PRINT:PRINT
1610 PRINT"ADJA MEG A MEGBIZHATOSAGI INTERVALLUM   FELSZELESSEGET!"
1620 INPUT R:IF R<=0 THEN PRINT"POZITIV SZAM!":GOTO 1620
1630 GOSUB 1800
1640 G=(K(L)*Q1/R)↑2
1650 GOTO 1860
1660 U=(M1-M2)/SQR(Q1↑2/N+Q2↑2/M)
1670 PRINT:PRINT"EGYOLDALI (1) VAGY KETOLDALI (2)        HIPOTEZISROL VAN SZO?"
1680 INPUT W
1690 IF W<>1 AND W<>2 THEN 1670
1700 IF W=1 THEN GOSUB 1730
1710 IF W=2 THEN GOSUB 1800
1720 GOTO 1860
1730 PRINT:PRINT"HANY SZAZALEKOS MEGBIZHATOSAGI SZINTET  VALASZT? (90-99.5)"
1740 PRINT"(0.5-ES PONTOSSAGGAL ADJA MEG!)"
1750 INPUT D  .
1760 IF D<90 OR D>99.5 THEN PRINT"NEM JO ADAT!":GOTO 1730
1770 IF D<>INT(D) AND D<>INT(D)+.5 THEN PRINT"0.5-ES PONTOSSAGGAL!":GOTO 1750
1780 L=D*2-179
1790 RETURN
1800 PRINT:PRINT"HANY SZAZALEKOS MEGBIZHATOSAGI SZINTET  VALASZT? (80-99)"
1810 INPUT D
1820 IF D<80 OR D>99 THEN PRINT"NEM JO ADAT!":GOTO 1800
1830 IF D<>INT(D) THEN PRINT"EGESZ SZAMOT ADJON MEG!":GOTO 1800
1840 L=D-79
1850 RETURN
```

4.4.3. Egymintás t-próba

Az előző pontban ismertetett u-próba abban az esetben alkalmazható, amikor a normális eloszlású valószínűségi változó szórása ismert és a várható értékre vonatkotó hipotézist szeretnénk ellenőrizni. Ha az elméleti szórás nem ismert, akkor a következő módon járunk el: Legyen a ξ valószínűségi változó eloszlása normális, vegyünk n elemű mintát és becsüljük az eloszlás elméleti szórását a korrigált tapasztalati szórással (s). Legyen az eloszlás ismeretlen várható értéke M. A

$$\mathrm{H}_0\colon\ M(\xi) = a_0$$

nullhipotézis vizsgálatára konstruáljuk a

$$t_{n-1} = \frac{\bar{x} - a_0}{\dfrac{s_n}{\sqrt{n}}}$$

statisztikát. Ha a nullhipotézis igaz, akkor a t_{n-1} statisztika eloszlása $n-1$ szabadsági fokú Student eloszlást követ.

Ha az ellenhipotézis az

$$M(\xi) \neq a_0$$

azaz mindegy számunkra, hogy kisebb-e a várható érték, mint az „előírt” a_0 vagy nagyobb, akkor ún. kétoldali próbát végzünk. A kétoldali, szimmetrikus kritikus tartomány:

$$X_k = \{t_{n-1} \leqq -t_\varepsilon \text{ vagy } t_{n-1} \geqq t_\varepsilon\}.$$

Adott ε szignifikanciaszinthez és $n-1$ szabadsági fokhoz tartozó t_ε kritikus értéket kiválasztva, ha a t_{n-1} statisztika értéke a $\{-t_\varepsilon, t_\varepsilon\}$ tartományon kívül esik (abszolút értéke nagyobb t_ε-nál), akkor a nullhipotézist elvetjük.

Ha az ellenhipotézis az

$$M(\xi) > a_0$$

vagy az

$$M(\xi) < a_0,$$

akkor csak az az eset jelenti a nullhipotézis elvetését, ha t a kritikus értéknél is nagyobb pozitív, vagy az ellentettjénél kisebb negatív érték (egyoldali próbák).

Kétoldali próbánál a megadott p szignifikanciaszintet úgy kell értelmezni, hogy a Student eloszlás mindkét végéből a $p/2$-nek megfelelő területet levágjuk, és az így kapott kritikus értékekkel hasonlítjuk össze a kiszámított t értéket.

Egyoldali próbánál a p szignifikanciaszint a Student eloszlás sűrűségfüggvényéből az egyik oldalon levágott terület mértéke.

Egymintás t próbára példaképpen válaszoljunk a következő kérdésekre: Egy cellulózüzemben a szulfitlúgból 9 mintát vettek, és mérték a szulfitlúg lúgosságát. A lúgossági számok a 9 mintára a következők voltak: 11,7; 12,2; 10,9; 11,4; 11,3; 12,0; 11,1; 10,7; 11,6.

A technológiai előírás 12,1-es lúgossági számot ír elő. Mondhatjuk-e, hogy a lúgossági szám megfelel-e a technológiai előírásoknak? Kijelentésünket 95%-os megbízható szinten kell megtennünk. A kérdés tulajdonképpen az, hogy a 9 elemű minta alapján mondhatjuk-e, hogy a lúgossági szám várható értéke 12,1? Mivel a szórás ismeretlen, azt is a mintából kell becsülnünk, így Student-próbát alkalmazunk.

Kiválasztjuk a STATISZTIKAI PRÓBÁK menüjéből az EGY- ÉS KÉTMINTÁS T-PRÓBA ÉS WELCH-PRÓBA menüjét, és azon belül az EGYMINTÁS T-PRÓBÁ-t. Adatainkat akár lemezről is bevihetjük, akár közvetlenül a billentyűzetről interaktív módon.

A MINTA ELEMEINEK SZÁMA: 9

Ezután rendre megadjuk a 9 lúgossági szám értéket, majd kiválasztjuk a T-ÉRTÉK KISZÁMÍTÁSÁ-t a képernyőn megjelenő menüből. Ez a program először a várható értéket kéri: a várható érték a technológiai előírásban szereplő 12.1. Miután mindkét irányú eltérés egyaránt káros, kétoldali próbát kell végezni. 95%-os megbízhatósági szintet választva a képernyőn a következő eredmény jeleníthető meg:

T-ÉRTÉK = −4,0406
95%-OS MEGBÍZHATÓSÁGI SZINTEN A HIPOTÉZIST EL KELL VETNI.
SZABADSÁGI FOK =8
KORR. TAPASZTALATI SZÓRÁSNÉGYZET=0,245

Valóban, erről úgy is meggyőződhetünk, ha kikeressük a Melléklet 2. táblázatá-

ból a 8 szabadsági fokú Student eloszlás értékét 0,05/2=0,025 szignifikancia-szinten:

$$t_{0,025,8} = 2{,}306.$$

Mivel a kiszámított t érték, −4,0406 kívül esik a {−2,306; 2,306} elfogadási tartományon, a nullhipotézist, azaz a lúgossági szám várható értékének a technológiai előírásban szereplő 12,1-del való egyezését el kell vetnünk.

4.4.4. Kétmintás t-próba és Welch-próba

Kétmintás t-próba

Ha két normális eloszlású valószínűségi változó várható értékeinek összehasonlítása a feladat és a szórásokat nem ismerjük, akkor a megfelelő t-próba csak abban az esetben alkalmazható, ha feltehetjük, hogy a két változó — előttünk ismeretlen — szórása megegyezik. A két szórás egyezését elméleti meggondolások igazolhatják vagy előzetes tapasztalataink indokolhatják. Ha ezt nem feltételezhetjük, akkor a szórások egyezéséről az F-próba alapján dönthetünk. Ha a szórások nem egyeznek meg, a t-próba helyett a Welch-próbával ellenőrizhetjük a várható értékek egyezésére vonatkozó hipotézisünket.

Legyen ξ és η normális eloszlású független valószínűségi változók $\sigma_1=\sigma_2$ szórással és vizsgáljuk meg a

$$H_0\colon\ M(\xi) = M(\eta)$$

nullhipotézist a ξ-re vett n elemű és az η-ra vett m elemű minták alapján. Jelöljük a ξ-re vett minta átlagát $\bar{x}$-szel, az η-ra vett minta átlagát $\bar{y}$-nal. Ha a nullhipotézis igaz, akkor a

$$t = \frac{\bar{x}-\bar{y}}{s\sqrt{1/n+1/m}}$$

valószínűségi változó $n+m-2$ paraméterű Student eloszlást követ, ahol s a különbség standard derivációja,

$$s^2 = \frac{(n-1)\,s_1^2+(m-1)\,s_2^2}{n+m-2}.$$

Az előbbi képleteket összevonva, átalakítás után a következő kifejezést kapjuk:

$$t_{n+m-2} = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{(n-1)\,s_1^2+(m-1)\,s_2^2}}\left(\frac{nm(n+m-2)}{n+m}\right)^{1/2}.$$

A kétoldali alternatíva esetén, tehát ha az ellenhipotézis

$$H_1\colon\ M(\xi) \neq M(\eta),$$

akkor $t_{n+m-2}>t_\varepsilon$ esetén a nullhipotézist elvetjük. t_ε az ε szignifikanciaszinthez és $n+m-2$ szabadsági fokhoz tartozó érték.

Ha ellenhipotézisünk

$$H_1'\colon\ M(\xi) > M(\eta),$$

akkor az alternatíva egyoldali, és a nullhipotézist

$$t_\varepsilon < t_{n+m-2} \quad \text{esetben vetjük el.}$$

Welch-próba

Ha a ξ és η független normális eloszlású valószínűségi változók, σ_1 és σ_2 szórásainak egyezését nem tételezhetjük fel, ill. az F-próba szerint $\sigma_1 \neq \sigma_2$, akkor a

$$H_\emptyset: M(\xi) = M(\eta)$$

nullhipotézis vizsgálatára a kétmintás t-próba nem alkalmazható, sőt ilyen esetben egzakt próba a kérdés vizsgálatára nem is adható. A kérdés eldöntésére Welch a következő statisztikát ajánlja:

$$t_f = \frac{\bar{x}_1 - \bar{x}_2}{\left(\frac{s_1^2}{n} + \frac{s_2^2}{m}\right)^{1/2}},$$

ahol $\bar{x}_1$, ill. $\bar{x}_2$ a ξ-re vett n elemű és η-ra vett m elemű minták átlaga, s_1^2 és s_2^2 a ξ és η tapasztalati szórásnégyzetei. A t_f statisztika a $H_\emptyset$ hipotézis fennállása esetén közelítőleg Student eloszlást követ f szabadsági fokkal, ahol f értéke a következő módon határozható meg:

Legyen

$$c = \frac{\frac{s_2^2}{m}}{\frac{s_1^2}{n} + \frac{s_2^2}{m}},$$

akkor

$$\frac{1}{f} = \frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}.$$

Ily módon ha f-et meghatároztuk, a t-eloszlás táblázatából a megfelelő szignifikanciaszinthez tartozó kritikus értékek leolvashatók. A próba használata (ellenhipotézisek, kritikus tartomány kijelölése) azonos a t-próbáéval. (A Student eloszlás kritikus értékeit a Melléklet 2. táblázata tartalmazza.)

A kétmintás t-próba és Welch-próba alkalmazása, beleértve a megfelelő programok használatát is, minden szempontból azonos, ezért használatukat egyetlen példán mutatjuk be.

Tegyük fel, hogy két szállítmány vasérc érkezett egy vaskohóba. A vasérc mangántartalma mindkét szállítmányban kismértékben szór, de a szórások lényegében megegyeznek. (A Fisher-próbával végzett vizsgálat is ezt mutatja 5%-os szignifikanciaszinten.) Mindkét szállítmányból 10 elemű mintát veszünk. A minták mangántartalmára (%) a következő értékeket mértük:

A szállítmány:	3,3	3,7	3,5	4,1	3,4	3,5	4,0	3,8	3,2	3,7
B szállítmány:	3,2	3,6	3,1	3,4	3,0	3,4	3,6	3,3	3,1	2,8

Kérdés: azonosnak tekinthető-e a szállítmányok mangántartalma, statisztikai értelemben van-e a mangántartalmak között szignifikáns eltérés 95%-os megbízhatósági szinten? Mi az a legkisebb szignifikanciaszint, amin a két szállítmány mangántartalma még éppen különbözőnek tekinthető? A kérdések megválaszolására a

STATISZTIKAI PRÓBÁK menüjéből az EGY- ÉS KÉTMINTÁS T-PRÓBA, valamint WELCH-PRÓBA programot választjuk ki. Mivel 2 mintánk van, de a szórások lényegében megegyeznek, ezért a KÉTMINTÁS T-PRÓBA alprogramot hívjuk. Az adatokat akár egy lemezen előkészített file-ból, akár billentyűzetről is bevihetjük. A program azt is megengedi, hogy ha már korábban kiszámítottuk az egyes minták átlagát és szórását, közvetlenül ezeket a „szekunder" adatokat adjuk meg, vagy a minták elemszámának megadása után rendre begépeljük az egyes minták mérési eredményeit. Ha a kiindulási adatokat megadtuk, a program megkérdezi, hogy egy- vagy kétoldali próbát kívánunk-e alkalmazni. Mivel mindegy számukra, melyik várható érték nagyobb, ezért természetesen kétoldali próbát kell alkalmaznunk (hiszen az ellenhipotézis a két szállítmány mangántartalmának, azaz valódi értékének különbözősége).

A megbízhatósági szintet 95%-ban rögzítve, a képernyőre a következő eredményt írhatjuk ki:

KÉTOLDALI HIPOTÉZIS ESETÉN
T-ÉRTÉK =2,987
95%-OS MEGBÍZHATÓSÁGI SZINTEN A HIPOTÉZIST EL KELL VETNI.
SZABADSÁGI FOK=18
KORR. TAPASZTALATI SZÓRÁSNÉGYZETEK
1. MINTÁNÁL: 0,0862
2. MINTÁNÁL: 0,0672

A program által adott döntést egyszerűen ellenőrizhetjük a Melléklet 2. táblázatának használatával. 5%-os szignifikanciaszinten a 18 szabadsági fokú Student eloszlás kritikus értéke 2,101. Mivel a számított t érték

$$t_{18} = 2{,}987 > t_{0.05,18} = 2{,}101,$$

a nullhipotézist, azaz a két minta mangántartalmainak azonosságát el kell vetni. A két mangántartalom 5%-os szignifikanciaszinten különbözőnek mondható. (Sőt, a táblázat adatainak vizsgálatával még az is kiderül, hogy még 0,5%-os szignifikanciaszinten is meg tudjuk különböztetni a két átlagértéket, mert $t_{0.005,18}=2.88$.)

4.4.5. Konfidencia intervallum kijelölése

Az u- és t-statisztika alkalmazásával kapcsolatban külön ki kell térnünk egy fontos fogalomra: a *konfidencia intervallum* fogalmára.

Eddigi tárgyalásunkból tudjuk, hogy a számtani átlag a várható érték becslésére vonatkozó statisztika, a szórást pedig a korrigált tapasztalati szórással (standard deviációval) becsülhetjük. Ezek a becslések egy-egy konkrét számot adnak, nem mondanak azonban semmit a kapott becslés bizonytalanságáról. Nem tudjuk tehát, hogy ha kiszámítunk egy átlagot egy 10 elemű minta alapján, akkor mennyire bízhatunk abban, hogy újabb 10 elemű mintát véve ugyanazt az átlagot kapjuk, ill nem tudhatjuk, hogy a régi átlaghoz várhatóan mennyire lesz közel az újonnan kiszámított. Ezért a számtani átlag ún. *pontbecslése* a várható értéknek, a standard deviáció pedig a szórásnak.

Statisztikai értelemben megalapozottabb lenne olyan becslést adni, hogy az illető kiszámított próbastatisztika a valódi értéknek milyen környezetébe kell hogy essen

az adott valószínűséggel. Ilyen becslés is lehetséges, az ilyen becslést *intervallumbecslésnek* nevezik a statisztikai irodalomban. Az ezzel kapcsolatos eljárás a *konfidencia intervallum kijelölése,* azé az intervallumé, amelybe a valódi értéknek (adott valószínűséggel) bele kell esnie. Konfidencia intervallumot mind az u-statisztika, mind a t-statisztika használatával kijelölhetünk. Ha a sokaság szórása ismert, az u-statisztikát (tulajdonképpen a standard normális eloszlás eloszlásfüggvényét), ha a sokaság szórását is a mintából becsültük, akkor a t-statisztikát (a Student eloszlás eloszlásfüggvényét) használjuk a konfidencia intervallum kijelölésére.

Ismert szórású, normális eloszlású sokaságból vett minta esetén ugyanis az egymintás u-statisztika képletét átrendezve a következő egyenlethez jutunk:

$$|\bar{X} - a_0| = u_\varepsilon \frac{\sigma}{\sqrt{n}},$$

azaz

$$a_0 \in \left\{\bar{X} \pm u_\varepsilon \frac{\sigma}{\sqrt{n}}\right\}.$$

Az utóbbi egyenlet szerint a sokaság várható értéke a mintaátlag $u \frac{\sigma}{\sqrt{n}}$ sugarú környezetébe kell, hogy essen az u érték számításánál felhasznált ε szignifikanciaszint esetén $1-\varepsilon$ megbízhatósággal.

Ha a sokaság szórása ismeretlen, és azt is a minta alapján becsültük, akkor az egymintás t-statisztika képletéből kell kiindulnunk a konfidencia intervallum kijelölésénél:

$$|\bar{X} - a_0| = t_{n-1,\varepsilon} \frac{S_n}{\sqrt{n}},$$

azaz

$$a_0 \in \left\{\bar{X} \pm t_{n-1} \frac{S_n}{\sqrt{n}}\right\},$$

vagyis a sokaság várható értéke $1-\varepsilon$ megbízhatósági szinten a mintaátlag $\pm t_{n-1} \frac{S_n}{\sqrt{n}}$ környezetébe kell, hogy essen.

Nézzünk most mind az u-, mind a t-statisztika alapján számított konfidencia intervallumra egy-egy példát.

a) A 4.4.1 pontban leírt példában egy folyóvíz higanytartalmára vonatkozó hipotézist ellenőriztünk. A minták átlagos higanytartalma 1,985 μg/l volt, a (valódi) szórás

$$\sigma = 0{,}04 \text{ g/l}.$$

Számítsuk most ki, hogy az eredeti feladat feltételei mellett a mintára kiszámított $\bar{x}=1{,}985$ érték milyen környezetébe esik a várható érték 95%-os megbízhatósággal!

Válasszuk tehát az EGYMINTÁS U-PRÓBA menüjéből a

MEGBÍZHATÓSÁGI INTERVALLUM KISZÁMÍTÁSA (M)

programágat, és adjuk meg ismét a 95%-ot mint megbízhatósági szintet. A képernyőn a következő eredmény jelenik meg:

A KONFIDENCIA INTERVALLUM 95%-OS MEGBÍZHATÓSÁGI SZINTEN

$$1{,}985 \pm 9{,}8 \cdot 10^{-3}.$$

Ez azt jelenti számunkra, hogy a várható érték 95%-os megbízhatósággal az 1,975—1,995 intervallumba esik. A 2,0-ás érték ezen kívül esik, ezért a nullhipotézis elutasítása mindenképpen jogos volt.

Fel kell azonban hívni a figyelmet arra, hogy ha ugyanezt az átlagértéket 28 mérés átlagaként kaptuk volna, akkor a konfidencia intervallum felső határa éppen 2 lenne, ezért a nullhipotézist el kellene fogadni. Erről könnyen meggyőződhetünk, többféle módon is. Egyik lehetőség, ha a MÉRÉSEK SZÁMÁNAK MEGHATÁROZÁSA (N) ágat választjuk, és konfidencia intervallum szélességeként 0,015-et adunk meg. Ekkor a képernyőn a következő eredmény jelenik meg: A MÉRÉSEK SZÁMA=28.

b) A 4.4.3 pontban az egymintás t-próbával kapcsolatban idéztünk egy példát, amelyben egy cellulózüzemben a szulfitlúg lúgossági számára vonatkozó nullhipotézist kellett ellenőriznünk. A minta átlaga 9 mérés alapján 11,43 volt, a korrigált tapasztalati szórásnégyzet pedig 0,245.

Kérdés, hogy mit mondhatunk most a lúgossági szám várható értékéről 95%-os megbízhatósággal? Erre a kérdésre is könnyen válaszolhatunk, ha most az egymintás t-próba menüjéből a MEGBÍZHATÓSÁGI INTERVALLUM MEGHATÁROZÁSA programágat választjuk. Itt újra meg lehet adni a megbízhatósági szintet (95%), és a képernyőn megjelenik a következő eredmény:

A KONFIDENCIA INTERVALLUM 95%-OS MEGBÍZHATÓSÁGI SZINTEN:

11,43 ±0,38

SZABADSÁGI FOK =8

KORR. TAPASZTALATI SZÓRÁSNÉGYZET=0,245

A válasz szerint a lúgossági szám várható értéke a {11,05, 11,81} intervallumba esik 95%-os valószínűséggel. Érdemes megjegyezni, hogy a párhuzamos mérések, amelyeket a pontosabb és megbízhatóbb eredmény érdekében végzünk, lényegében a Student eloszlás alkalmazásával végzett konfidencia intervallum kijelölésre vezetnek. Ha valaki egy uszoda vizének klórtartalmát vagy egy mezőgazdasági termék hozamát 5-ször egymás után megméri, akkor ezt nem azért teszi, hogy az illető 5 elemű mintára kelljen szorítkoznia megállapításának, hanem a mérési eredmény átlagának és szórásának megadásával az eredeti jelenséget (az uszoda vizének klórtartalmát vagy a termékhozamot) szeretné jellemezni. Az ismételt mérésekből származó eredmény korrekt megadási módja ezért a következő: Legyen a mérendő változó x, tegyük fel, hogy kiszámítottuk a mérések alapján számtani átlagát ($\bar{x}$) és szórását (s). Ekkor a mérési eredmény megadásának korrekt módja:

$$\bar{x} \pm t_{(n-1,\varepsilon)} \frac{s}{\sqrt{n}},$$

ahol $t_{(n-1,\varepsilon)}$ az ε szignifikanciaszinthez és $n-1$ szabadsági fokhoz tartozó kritikus érték. (Itt természetesen mindig kétoldali próbáról, ezért kétoldali szimmetrikus tartományról van szó.)

Végül megemlítendő, hogy mivel az átlag szórása $s/\sqrt{n}$, ez az érték a minta

elemszámának (a párhuzamos mérések számának) növelésével csökkenthető. Ezért érdemes foglalkozni azzal a kérdéssel, hogy adott terjedelmű megbízhatósági intervallumhoz milyen elemszámú minta (hány párhuzamos mérés) tartozik.

T-STATISZTIKA

A változók listája:

N, M.	— mintacsoportok elemszáma,
X(I), Y(I)	— mintacsoportok elemei,
S1, S2	— elemek összegzése,
M1, M2	— elemek átlaga,
Q1, Q2	— korrigált tapasztalati szórásnégyzetek,
P	— várható érték,
R1, R2, O, C, F, H, W(D)	— részeredmények,
T	— t-érték,
D	— szabadsági fok
A	— konfidencia intervallum (számított),
R	— konfidencia intervallum (megadott),
G	— mérések száma,
L	— segédváltozó,
B	— megbízhatósági szint,
K(D, L)	— t kritikus értékei,
Q	— egy- vagy kétoldali hipotézisnél választás,
OX, OY	— adatmátrixból kijelölt oszlopok száma,
D$	— menünél választás,
A$	— Igen-Nem választás,
B$	— egymintás t-próba menüjénél választás,
ID	— kiíratásnál választás (képernyő vagy nyomtató),
N$	— adatfile neve,
HS, HN$, S, SZ	— lemezhiba ellenőrzése.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—240:	A program.
240—410:	DATA: t kritikus értékei.
420:	T kritikus értékeinek beolvasása.
430—570:	Menű kiírása, választás, elágazások.
580—1420:	Adatbevitel.
1430—1670:	Egymintás t-próba almenüjének kiírása, választás, elágazások.
1680—1710:	T-érték kiszámítása.
1720—1750:	Megbízhatósági intervallum kiszámítása.
1760—1860:	Mérések számának meghatározása.
1870—1890:	Kétmintás t-próba számítása.
1900—1930:	Welch-próba számítása.
1940—1990:	Egyoldali vagy kétoldali hipotézisre választás, elágazás.
2000—2100:	Szubrutin megbízhatósági szint választására kétoldali hipotézis esetén.
2110—2210:	Szubrutin megbízhatósági szint választására egyoldali hipotézis esetén.

2220—2740: Eredmények kiírása választás szerint képernyőre vagy printerre.
2750—3050: A program által nyújtott lehetőségek rövid ismertetése.
3060: Visszatérés a FŐMENÜ c. programra.
3070—3540: Adatbeolvasás lemezről és a vizsgálandó mintacsoportok kijelölése. Közben lemezhiba ellenőrzése a 3550-es szubrutinnal.
3550—3660: Lemezhibát ellenőrző szubrutin.
3670—3770: Várakozó szubrutin.
3780—3850: Várakozó szubrutin Igen-Nem válaszra.

```
1680 INPUT"ADJA MEG A VARHATO ERTEKET ";P
1690 T=1E37*SGN(M1-P):IF Q1<>0 THEN T=(M1-P)/SQR(Q1/N)
1700 D=N-1:DD=D:IF D>31 THEN D=31
1710 GOTO 1940
1720 GOSUB 2110
1730 D=N-1:DD=D:IF D>31 THEN D=31
1740 A=K(D,L)*SQR(Q1/N)
1750 GOTO 2220
1760 PRINT:PRINT
1770 PRINT"ADJA MEG A KONFIDENCIA INTERVALLUM       FELSZELESSEGET!"
1780 INPUT R:IF R<=0 THEN PRINT"POZITIV SZAM!":GOTO 1780
1790 GOSUB 2110:H=SGN(R)*1E37:IF Q1<>0THEN H=R/SQR(Q1)
1800 FOR D=1 TO 31:W(D)=K(D,L)/SQR(D+1):NEXT D
1810 IF H>W(1) THEN G=1:GOTO 2220
1820 IF H<W(31) THEN G=33:GOTO 2220
1830 FOR D=1 TO 31:IF H>=W(D) THEN G=D+1:D=31
1840 NEXT D:GOTO 2220
1850 G=(K(L)*Q1/R)↑2
1860 GOTO 2220
1870 U=SGN(M1-M2)*1E37:IF Q1<>0 OR Q2<>0 THEN U=(M1-M2)/SQR((N-1)*Q1+(M-1)*Q2)
1880 T=U:IF ABS(U)<1E37 THEN T=U*SQR(N*M*(N+M-2)/(N+M))
1890 D=N+M-2:DD=D:IF D>31 THEN D=31
1895 GOTO 1940
1900 T=SGN(M1-M2)*1E37:IF Q1<>0 OR Q2<>0 THEN T=(M1-M2)/SQR(Q1/N+Q2/M)
1910 C=0:IF Q1<>0 OR Q2<>0 THEN C=Q2/M/(Q1/N+Q2/M)
1920 F=C*C/(M-1)+(1-C)*(1-C)/(N-1)
1930 D=1/F:DD=D:IF D>31 THEN D=31
1940 PRINT:PRINT"EGYOLDALI (1) VAGY KETOLDALI (2)        HIPOTEZISROL VAN SZO?"
1950 INPUT Q
1960 IF Q<>1 AND Q<>2 THEN 1940
1970 IF Q=1 THEN GOSUB 2000
1980 IF Q=2 THEN GOSUB 2110
1990 GOTO 2220
2000 PRINT:PRINT"HANY SZAZALEKOS MEGBIZHATOSAGI SZINTET  VALASZT?"
2010 PRINT"A LEHETOSEGEK:90,95,97.5,99,99.5,99.95"
2020 INPUT B
2030 IF B<>90 AND B<>95 AND B<>97.5 AND B<>99 AND B<>99.5 AND B<>99.95 THEN 2020
2040 IF B=90 THEN L=1
2050 IF B=95 THEN L=2
2060 IF B=97.5 THEN L=3
2070 IF B=99 THEN L=4
2080 IF B=99.5 THEN L=5
2090 IF B=99.95 THEN L=6
2100 RETURN
2110 PRINT:PRINT"HANY SZAZALEKOS MEGBIZHATOSAGI SZINTET  VALASZT?"
2120 PRINT"A LEHETOSEGEK:80, 90, 95, 98, 99, 99.9"
2130 INPUT B
2140 IF B<>80 AND B<>90 AND B<>95 AND B<>98 AND B<>99 AND B<>99.9 THEN 2130
2150 IF B=80 THEN L=1
2160 IF B=90 THEN L=2
2170 IF B=95 THEN L=3
2180 IF B=98 THEN L=4
2190 IF B=99 THEN L=5
2200 IF B=99.9 THEN L=6
2210 RETURN
```

4.5. Szórások összehasonlítása

4.5.1. Két szórás összevetése: F-próba (Fisher-próba)

Két független, normális eloszlású valószínűségi változó szórásának összehasonlítására alkalmas. A nullhipotézis itt arra vonatkozik, hogy a két normális eloszlású ismeretlen várható értékű sokaság szórása azonos, azaz:

$$H_\emptyset: \sigma_1 = \sigma_2,$$

ahol σ_1 az egyik, σ_2 a másik sokaság szórása. Ha $H_\emptyset$ igaz, akkor a két sokaságból vett n, ill. m elemű mintából s_1, ill. s_2 korrigált tapasztalati szórásnégyzetek hányadosa F eloszlást követ, $n-1$ és $m-1$ szabadságfokkal:

$$F = \frac{s_1^2}{s_2^2}.$$

Ezt a valószínűségi változót használjuk a nullhipotézis ellenőrzésére. Megjegyezzük, hogy az F értékek kiszámításánál mindig a nagyobb szórásnégyzet kerül a számlálóba. Ez egyszerűsíti a próba elvégzését és a kritikus értékek táblázatos megadását; tulajdonképpen ekkor mindig a

$$H_1: \sigma_1^2 > \sigma_2^2$$

ellenhipotézissel szemben vizsgáljuk a nullhipotézis érvényességét, azaz mindig egyoldali próbát végezhetünk. Ha az s_1^2/s_2^2 hányados jóval nagyobb, mint 1, nyilván valószínűtlen, hogy a nullhipotézis igaz lenne.

Az F-eloszlás kritikus értékeit a Melléklet 4. táblázata tartalmazza. Ezek a kritikus értékek a szignifikanciaszinten kívül a számláló és a nevező szabadsági fokától ($n-1$, ill. $m-1$ értékétől) is függenek. Ez „3-dimenziós" táblázatot jelentene, ezért egy-egy táblázat egy adott szignifikanciaszinthez tartozik.

Nem nehéz belátni, hogy ha mind a számláló, mind a nevező szabadsági foka kicsi, már „megengedhető" az, hogy „véletlenül" más szóráshányadost kapjunk. Ez azt jelenti, hogy adott szignifikanciaszintnél a szabadságfokok csökkenésével nő a táblázatban szereplő F_{krit} értéke.

A próba használatával kapcsolatban megjegyezzük, hogy ha a kisebb szórást írnánk a számlálóba, akkor a táblázatos értékeket úgy kellene használni, hogy az előírt értéknél kisebb szóráshányados esetén kellene a nullhipotézist elutasítani. Ilyenkor az

$$F(1-\varepsilon, n-1, m-1) = \frac{1}{F} \quad (\varepsilon, m-1, n-1)$$

számítási formulával kellene a kritikus értékeket átszámítani.

A próba használatának bemutatására tekintsük ismét a kétmintás t-próbánál szereplő példát: két független, 10 elemű mintánk volt, $s_1^2=0{,}0862$ és $s_2^2=0{,}0672$ szórásnégyzettel. A szórások összehasonlítása F-próbával csupán egy hányados kiszámítását és a hányadosnak a táblázatos értékkel való összehasonlítását igényli. Használjuk most mégis az F-próba elvégzésére írott programot, mert a program többet nyújt a próba egyszerű elvégzésénél. Hívjuk be a STATISZTIKAI PRÓBÁK

menüjéből az F-PRÓBA ÉS BARTLETT-PRÓBA programot. Ekkor a következő almenü jelenik meg a képernyőn:

F-PRÓBA ÉS BARTLETT-PRÓBA
AZ ALÁBBI LEHETŐSÉGEK KÖZÜL VÁLASZTHAT:
F-PRÓBA (F)
BARTLETT-PRÓBA (B)
ISMERTETÉST KÉREK (I)
VÉGE (V)

Az F-PRÓBA menüágat kiválasztva, az adatokat lemezen korábban előkészített file-ból vagy billentyűzetről is bevihetjük, akár a teljes adatsor, akár a már korábban kiszámított szórások megadásával.

A minták elemszámát és szórásnégyzetét megadva a következő eredmény írható ki a képernyőre (vagy a nyomtatóra, a felhasználó választása szerint).

F-ÉRTÉK =1,283
SZÁMLÁLÓ SZABADSÁGI FOKA=0
NEVEZŐ SZABADSÁGI FOKA =9
A SZIGNIFIKANCIASZINT = ,358

Mit is jelent az utolsó sorban szereplő 0,358-as szignifikanciaszint? Azt jelenti, hogy csak akkor utasíthatnánk el a nullhipotézist, azaz a két szórás azonosságát, ha 35,8%-os kockázatot vállalnánk az igaz hipotézis elutasítására. A szokásos 5%-os szignifikanciaszintnél ezt nem tehetnénk meg, tehát 5%-os szignifikanciaszinten (azaz 95%-os megbízhatósági szinten) a nullhipotézist el kell fogadnunk. A két szórás azonosnak tekinthető. Az 5%-os szignifikanciaszinthez tartozó kritikus érték 3,18. Mivel $F_{n-1,m-1}<F_{\text{krit}}$, a $H_\emptyset$-t elfogadjuk.

4.5.2. Több szórás egyezésének vizsgálata: Bartlett-próba

Két normális eloszlású valószínűségi változó szórása egyezésének eldöntésére megismertük az F-próbát. Több normális eloszlású valószínűségi változó szórásának egyezését egy lépésben a Bartlett-ről elnevezett próbával vizsgálhatjuk. Nullhipotézisünk a következő:

$$H_\emptyset:\ \sigma(\xi^1)=\sigma(\xi^2)=\ldots=\sigma(\xi^k),$$

ahol $\xi^1, \xi^2, \ldots, \xi^k$ k db normális eloszlású valószínűségi változó. Tekintsük az egyes valószínűségi változókra vett $n_1, n_2, \ldots, n_k$ elemű mintákat, és legyenek a mintákból számított korrigált tapasztalati szórásnégyzetek rendre;

$$s_{n_1}^2, s_{n_2}^2, s_{n_3}^2, \ldots, s_{n_k}^2.$$

Vezessük be az $f_i=n_i-1$ $(i=1, 2, \ldots, k)$ és az $f=\sum_{i=1}^{k} f_i$ jelöléseket. Az f_i értékekkel súlyozott összesített szórás:

$$s_n^2=\frac{1}{f}\sum_{i=1}^{k} f_i s_{n_i}^2.$$

Bartlett kimutatta, hogy a

$$K^2=\frac{2{,}3026}{c}\left(f\lg s_n^2-\sum_{i=1}^{k} f_i\lg s_{n_i}^2\right)$$

valószínűségi változó közelítőleg χ^2 eloszlást követ $k-1$ paraméterrel. A képletben a lg a 10-es alapú logaritmust jelöli, míg

$$c = 1+\frac{1}{3(k-1)}\left\{\sum_{i=1}^{k}\frac{1}{f_i}-\frac{1}{f}\right\}.$$

A közelítés akkor jó, ha minden i-re $n_i \geqq 4$ $i=1, 2, \ldots, k$. A nullhipotézist $(1-\varepsilon)$ megbízhatósági szinten akkor fogadjuk el, ha a χ^2 próba táblázatából kikeresve a $\chi^2(k-1, 1-\varepsilon)$ kritikus értéket, a $K^2 < \chi^2$ reláció teljesül. A próbát leggyakrabban a szórásanalízis előkészítő lépéseként alkalmazzák. A próba elvégzésére készített program illusztrálására tekintsük a következő példát:

Egy műtrágya nitrogéntartalma 20%. 5 egymást követő napon 12—12 zsák műtrágyát kiválasztva, elemezték a műtrágyát. A műtrágyák átlagos nitrogéntartalma és a nitrogéntartalom szórása az 5 db 12 elemű sorozatnál a következőképpen alakult:

i	1.	2.	3.	4.	5.
$\bar{x}_i$	18,73	21,20	19,73	20,34	19,81
S_i^2	3,06	1,97	3,2	5,05	2,98

A várható értékek összehasonlításáról a szórásanalízisben bővebben lesz szó. Elöljáróban csak annyit, hogy a szórásanalízis alkalmazásához igazolnunk kell, hogy az egyes minták szórásai azonosak. Válaszoljunk tehát most arra a kérdésre, hogy a különböző napokon gyártott műtrágyák nitrogéntartalmának szórása megegyező-e?

Hívjuk be tehát a STATISZTIKAI PRÓBÁK menüjéből az F-PRÓBA ÉS BARTLETT-PRÓBA programot és válasszuk a képernyőn megjelenő almenüből a BARTLETT-PRÓBA programját. A képernyőn megjelenő kérdésekre válaszolva gépeljük be a minták számát (5), az egyes minták elemszámát (12) és számított szórásnégyzeteiket. Válaszképpen a képernyőn (vagy a felhasználó kívánsága szerint a nyomtatón) a következő eredmény jelenik meg:

KHI-NÉGYZET =2,4069
SZABADSÁGI FOK=4
A VALÓSZÍNŰSÉG=0,3386

A kiírt valószínűség azt jelenti, hogy a nullhipotézist, azaz a szórások azonosságát csak 33,86%-os szignifikanciaszinten vethetnénk el. A szokásos 5%-os szignifikanciaszinten a Melléklet 3. táblázata szerint a χ^2 kritikus értéke 9,488. Mivel a számított khi-négyzet érték, $2{,}4069 < 9{,}488 = \chi^2_{\text{krit}}$, a szórások azonosságát elfogadjuk.

F-PRÓBA ÉS BARTLETT-PRÓBA

A változók listája:

L	— vizsgálandó mintacsoportok száma,
N(I)	— a mintacsoportok elemszáma,
X(I, J)	— a mintacsoportok elemei,
S(I)	— a mintacsoportok elemeinek összegzése,
G(I)	— f-próbánál a szabadsági fok,
M(I)	— a mintacsoportok elemeinek átlaga,
Q(I)	— korrigált tapasztalati szórásnégyzetek,

Y	— f-érték,
K	— khi-négyzet értéke,
F, R(I), P, D, W, H, C	— részeredmények,
PP	— f-próbánál a szignifikanciaszint,
A, B, F1, A1, B1, Z	— f-eloszlás számításánál a változók,
J, A, B, S, T, WW	— khi-négyzet eloszlás számításánál a változók,
M, N	— lemezről beolvasott adatmátrix oszlopainak és sorainak száma,
OX	— adatmátrixból kijelölt oszlop száma,
D$	— menünél választás,
A$	— Igen-Nem választás,
ID	— kiíratásnál választás (képernyő vagy nyomtató),
N$	— adatfile neve,
HS, HN$, S, SZ	— lemezhiba ellenőrzése.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—240:	A program neve.
250—390:	Menü kiírása, elágazások.
400—660:	Adatbevitel billentyűzetről, korrigált tapasztalati szórásnégyzet kiszámítása.
670—1040:	Adatmátrix kiíratása.
1050—1140:	F-érték kiszámítása.
1140—1350:	Az f-értékhez tartozó valószínűség kiszámítása.
1360—1450:	Khi-négyzet kiszámítása.
1460—1610:	A khi-négyzethez tartozó valószínűség kiszámítása.
1620—1960:	Eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
1970—2180:	A program által nyújtott lehetőségek rövid ismertetése.
2190:	Visszatérés a FŐMENÜ c. programra.
2200—2620:	Adatbeolvasás lemezről és a korrigált tapasztalati szórásnégyzet kiszámítása. Közben lemezhiba ellenőrzése a 2630-as szubrutinnal.
2630—2740:	Lemezhibát ellenőrző szubrutin.
2750—2850:	Várakozó szubrutin.
2860—2930:	Várakozó szubrutin Igen-Nem válaszra.

```
470 REM******************************
480 REM*                            *
490 REM* ADATBEVITEL BILLENTYUZETROL*
500 REM*                            *
510 REM******************************
520 PRINT:PRINT"A TELJES ADATSORT KIVANJA BEVINNI     (I)"
530 PRINT"VAGY MAR KISZAMITOTTA A SZORASO(KA)T (N)":GOSUB 2910:C$=A$
540 PRINT
550 PRINT
560 F=0
570 FOR I=1 TO L:S(I)=0
580 PRINT:PRINTI;". MINTA ELEMSZAMA=";:INPUT N
590 IF N<2 OR N>200 OR N<>INT(N) THEN PRINT"2 ES 200 KOZTI EGESZ!":GOTO 580
600 N(I)=N:G(I)=N-1:F=F+G(I):PRINT
610 IF C$="N" THEN INPUT"A KORR. TAP. SZORASNEGYZET";Q(I):GOTO 660
620 FOR J=1 TO N
630 PRINT"X(";J;",";I;")=";:INPUT X(J,I)
640 S(I)=S(I)+X(J,I):NEXT J:M(I)=S(I)/N(I):R(I)=0
650 FOR J=1 TO N(I):R(I)=R(I)+(X(J,I)-M(I))↑2:NEXT J:Q(I)=R(I)/G(I)
660 NEXT I
670 REM******************************
680 REM*                            *
690 REM* AZ ADATOK KIIRASA          *
700 REM*                            *
710 REM******************************
720 PRINT:PRINT"KIIRJUK AZ ADATOKAT";:GOSUB 2910:PRINT
730 IF A$="N" THEN 1090
740 PRINT"KEPERNYO(3) VAGY PRINTER(4) ";:INPUT ID
750 IF ID<>3 AND ID<>4 THEN 740
760 FOR O=3 TO ID:OPENO,O:IFO=3 THEN PRINT CHR$(147)
770 X$="F":IF D$="B" THEN X$="BARTLETT"
780 PRINT#O,"         ";X$;"-PROBA"
790 PRINT#O
800 PRINT#O:PRINT#O,"AZ ADATSOROK"
810 IF C$="N" THEN PRINT#O,"  ISMERETLENEK":GOTO 970
820 NX=N(1):FOR I=2 TO L:IF N(I)>NX THEN NX=N(I)
830 NEXT I
840 SN=3:I1=1:K1=0:IF O=4 THEN SN=6
850 I2=I1+SN-1:IF L<I2 THEN I2=L
860 PRINT#O:PRINT#O," I ";:FOR J=I1 TO I2
870 PRINT#O,RIGHT$(SP$+"X"+MID$(STR$(J),2)+" ",12);:NEXT J:PRINT#O:PRINT#O
880 FOR I=1 TO NX:K1=K1+1:IF K1<15 OR O=4 OR I=NX THEN 920
890 PRINT
900 PRINT"KERI A TOBBI ADATOT IS";:GOSUB 2910:IF A$="N" THEN I=NX:GOTO 950
910 K1=0:PRINT CHR$(147)
920 PRINT#O,RIGHT$(SP$+STR$(I),3);:FOR J=I1 TO I2
930 X$=LEFT$(SP$,12):IF I<=N(J) THEN X$=RIGHT$(SP$+STR$(X(I,J)),12)
940 PRINT#O,X$;:NEXT J:PRINT#O
950 NEXT I:IF I2<>L THEN I1=I1+SN:GOTO 850
960 GOSUB 2810
970 IF O=3 AND C$="I" THEN PRINT CHR$(147)
980 PRINT#O:PRINT#O,"ADATSOR   ADATSZAM    KORR.TAP."
990 PRINT#O:PRINT#O,"                      SZORAS↑2":PRINT#O
1000 FOR I=1 TO L:PRINT#O,RIGHT$(SP$+STR$(I),5);"          ";
1010 PRINT#O,RIGHT$(SP$+STR$(N(I)),3);"     ";RIGHT$(SP$+STR$(Q(I)),12)
1020 NEXT I
1030 PRINT#O:IF O=3 THEN GOSUB 2810
1040 CLOSE O:NEXT O
1050 REM******************************
1060 REM*                            *
1070 REM*     SZAMITASOK             *
1080 REM*                            *
1090 REM******************************
1100 PRINT CHR$(147)
1110 IF D$="B" THEN 1360
1120 IF Q(1)>=Q(2) THEN Y=Q(1)/Q(2):GOTO 1150
```

```
1130 Y=Q(2)/Q(1)
1140 V=G(1):G(1)=G(2):G(2)=V
1150 PP=1
1160 IF Y<1 THEN 1210
1170 A=G(1)
1180 B=G(2)
1190 F1=Y
1200 GOTO 1240
1210 A=G(2)
1220 B=G(1)
1230 F1=1/Y
1240 A1=2/9/A
1250 B1=2/9/B
1260 Z=ABS((1-B1)*F1↑(1/3)-1+A1)/SQR(B1*F1↑(2/3)+A1)
1270 IF B<4 THEN 1300
1280 PP=.5/(1+Z*(.196854+Z*(.115194+Z*(.000344+Z*.019527))))↑4
1290 GOTO 1320
1300 Z=Z*(1+.08*Z↑4/B↑3)
1310 GOTO 1280
1320 IF Y<1 THEN 1340
1330 GOTO 1670
1340 PP=1-PP
1350 GOTO 1670
1360 P=0:D=0:W=0
1370 DEF FN LG(X)=LOG(X)/LOG(10)
1380 FOR I=1 TO L
1390 P=P+G(I)*Q(I)
1400 D=D+G(I)*FNLG(Q(I))
1410 W=W+1/G(I)-1/F
1420 NEXT I
1430 H=P/F
1440 C=1+W/(3*(L-1))
1450 K=2.3026/C*(F*FNLG(H)-D)
1460 J=1
1470 FOR I=(L-1) TO 2 STEP -2
1480 J=J*I
1490 NEXT I
1500 A=K↑(INT(L/2))*EXP(-K/2)/J
1510 IF INT((L-1)/2)=(L-1)/2 THEN 1540
1520 B=1E37:IF K<>0 THEN B=SQR(2/K/3.1415927)
1530 GOTO 1550
1540 B=1
1550 S=1:T=1
1560 WW=L-1
1570 WW=WW+2
1580 T=T*K/WW
1590 IF T<.00001 THEN 1670
1600 S=S+T
1610 GOTO 1570
```

5. SZÓRÁSELEMZÉS

A szóráselemzés alapgondolata R. A. Fisher nevéhez fűződik. Segítségével eldönthető, hogy egy vagy több adott tényező (ún. faktor) gyakorol-e ténylegesen hatást valamely kísérlet eredményére (összehasonlító kísérletezés).

Néhány probléma, amely a szóráselemzés segítségével megoldható:

a) Több szennyvíztisztítási technológiával azonos minőségű szennyvizet tisztítunk. Tudni szeretnénk, hogy ezen technológiák által kibocsátott szennyvíz valamely paramétere ugyanazon érték körül ingadozik-e.

b) Háromféle korrózióvédő festéket kell összehasonlítanunk korrózióállóság szempontjából, háromféle minőségű acélra felhordva. A korrózióállóság megítélésénél azonban szeretnénk tisztán látni, hogy a jobb korrózióállóság a jobb acélminőség vagy az eltérő festéktípus következménye.

c) Különböző kukoricavetőmag-mintákat kell összehasonlítanunk egy termelőszövetkezetben. A szövetkezet földjei azonban eltérő adottságúak.

Előre kell bocsátani, hogy a szóráselemzésben a faktorok hatását egy folytonos valószínűségi változóra vizsgáljuk, s kikötjük, hogy a fellépő valószínűségi változók normális eloszlásúak és azonos szórásúak legyenek.

Példánkban több azonos szórású, normális eloszlású valószínűségi változó várható értékének összehasonlítása a cél. Két ilyen változó összehasonlítására a t-próbát már ismerjük. Most ennek több változóra való kiterjesztését tárgyaljuk. Megjegyezzük, hogy minden lehetséges pár t-próbával történő összehasonlítása nem lenne célszerű, mert a döntés szignifikanciaszintje függ az összehasonlítandó kísérletek számától. Gondoljuk meg, hogy 5%-os szignifikanciaszinten, 7 kísérlet várható értékének összehasonlítása esetén 21 t-próbát kellene alkalmazni, és annak valószínűsége, hogy H_0 igaz, de véletlenül valamelyik páros összehasonlítás mégis szignifikáns eltérést mutat, $1-(1-0{,}05)^{21}=0{,}66$ (66%!).

A második feladatban a változó várható értékét kétféle hatás befolyásolja: a bevonat és az acélminőség. Kérdés, hogy valóban léteznek-e ilyen hatások vagy sem. Hasonló kérdést vet fel a harmadik példa is.

A továbbiakban feltételezzük, hogy mindig véletlen kísérletezési sorrenddel kaptuk adatainkat. Ez nagyon fontos lehet az ún. módszeres hibák elkerülésében. (Pl. a kísérleti eredmény a műszerek kalibrációjának fokozatos eltolódása miatt szisztematikusan eltolódhat valamilyen irányban, az időjárás hatása a kísérlet eredményére elfedheti a vizsgálni kívánt beavatkozás hatását stb.) Feltételezzük azt is, hogy a kísérletező tisztában van a mérések ismételhetőségével és az ismétlések fontosságával is.

A kísérleti eredmények összehasonlítása és az egyes faktorok kísérleti eredményekre

gyakorolt hatásának különválasztására gondolatban felbontjuk az egyes észlelt eredményeket a következőképpen:

$$x_{ij} = \mu + t_i + \varepsilon_{ij} \quad (i = 1, 2, \ldots, c;\ j = 1, \ldots, n_i),$$

ahol x_{ij} a kísérlet eredménye a j-edik mérési sorozatban az i-edik kezelés mellett; c az összehasonlítandó kísérletsorozatok (kezelések) száma (pl. az első példában a különböző szennyvíztisztítási technológiák száma); n_i a mérések száma egy-egy kísérletsorozatban (kezelés során); μ a kezelés nélküli kísérlet várható értéke; t_i az i-edik kezelés hatása a várható értékre; ε_{ij} véletlen hiba.

A véletlen hiba következtében fellépő szórást *maradék szórásnak* nevezi a szakirodalom. A maradék elnevezés értelmezése az előbbi additív felbontásból nyilvánvaló: a külön-külön értelmezett individuális hatások különválasztása után megmaradó, megmagyarázatlan szóródás.

Természetesen az egyes kísérletsorozatok (kezelések) kimenetelének várható értékével is felírhatjuk a modellt:

$$x_{ij} = \mu_i + \varepsilon_{ij}$$

és ez esetben

$$\mu_i = \mu + t_i, \quad \sum_{i=1}^{c} t_i = 0.$$

A várható értéket $\bar{x}$-sal, a μ_i várható értéket $\bar{x}_i$-vel becsülhetjük, és így az i-edik kezelés t_i hatásának becslése: $\hat{t}_i = \bar{x}_i - \bar{x}$. Az új fogalmak gyakorlására tekintsük a következő példát: Négy különböző gyár rugóacélt gyárt. Nevezzük a gyárakat A, B, C és D gyárnak. Az acél szakítószilárdságának összehasonlítására 10—10 mintát kértünk mindegyik gyártól, véletlenszerűen kiválasztva a mintákat egy-egy nagyobb szállítmányból. A kapott szakítószilárdság-értékek dimenziómentes alakra hozva a következők voltak:

Gyár	A	B	C	D
Szakítószilárdság	55	70	70	90
	50	80	60	115
	80	85	65	80
	60	105	75	70
	70	65	90	95
	75	100	45	100
	40	90	95	105
	45	95	70	90
	80	100	65	100
	70	70	75	60
Csoportátlag, $\bar{x}_i$	62,5	86,0	70,5	90,5
Szórásnégyzet, s_i^2	212,5	184	253,8	274,7
$\hat{t}_i = \bar{x}_i - \bar{x}$	$-14{,}9$	8,6	$-6{,}9$	13,1

$\bar{x} = 77{,}4.$

A feladat tehát a H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ nullhipotézis, vagy ami ezzel egyenértékű, a H_0: $t_i = 0$ minden i-re nullhipotézis vizsgálata azzal a feltételezéssel, hogy a mérési eredmények normális eloszlásúak és az egyes kísérletsorozatokban (az egyes kezelé-

sek során) a szórások azonosak, azaz $\sigma_1^2=\sigma_2^2=\sigma_3^2=\sigma_4^2$. Ez utóbbi feltételezést a Bartlett-próbával ellenőrizhetjük (l. a Statisztikai próbák c. fejezetet).

5.1. Szóráselemzés egyszeres osztályozással

Ebben a fejezetben az előző pontban leírt nullhipotézis (ti., hogy az egyes kísérletsorozatok eredménye lényegében azonos, azaz az egyes kezelések nem befolyásolják a kísérletek várható eredményét) vizsgálatára mutatunk be egy módszert. A módszer lényege az, hogy az összes megfigyelésre együttesen kiszámított szórást két összetevőre bontjuk:

— az egyes kísérletsorozatok átlagértékei ($\bar{x}_i$) és a teljes átlag ($\bar{x}$) eltéréséből számított szórásra (ún. csoportok közötti szórás) és a

— a csoportokon belüli szórásra.

Az így értelmezett két szórás hányadosát F-próbával hasonlítjuk össze. (A tárgyalásban nem térünk ki a módszer mélyebb matematikai-statisztikai alapjaira, de megjegyezzük, hogy a módszer a Fisher—Cochran-féle szórásfelbontási tételen alapul.)

Tekintsük először az i-edik kezelés mellett kapott mérési eredmények korrigált tapasztalati szórásnégyzetét. Ez definíció szerint

$$s_i^2 = \sum_{j=1}^{n_i} \frac{(x_{ij}-\bar{x}_i)^2}{n_i-1}, \quad i = 1, 2, \ldots, c,$$

ahol c a csoportok (kezelések) száma; n_i az ismételt megfigyelések száma az egyes csoportokban. Vezessük be az $n=\sum_{i=1}^{c} n_i$ jelölést a megfigyelések teljes számára. Ha H_0 igaz, minden s_i^2 ugyanannak a közös σ^2-nek a becslése. Az összes megfigyelésre együttesen számított szórásnégyzet kiszámítható a csoportokon belüli varianciákból:

$$s^2 = \sum_{i=1}^{c} \frac{s_i^2}{c} = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n_i} \frac{(x_{ij}-\bar{x}_i)^2}{(n_i-1)}}{c} = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2}{n-c}.$$

Ez az s^2 mindenképpen becslése a közös σ^2-nek, akár igaz a nullhipotézis, akár nem. A csoportok közötti szórásnégyzet definíció szerint

$$\sum_{i=1}^{c} \frac{(\bar{x}_i-\bar{x})^2}{c-1}.$$

Mivel a csoportátlagok szórása $\sigma/\sqrt{n}$, az előbbi szórásnégyzet σ^2/n becslése, hacsak a nullhipotézis igaz. Ezért (a nullhipotézis fennállása esetén) σ^2-re becslést kaphatunk a csoportok közötti szórásnégyzet n_i-szeresével is. Jelöljük ezt s_c^2-vel,

$$s_c^2 = \sum_{i=1}^{c} \frac{n_i(\bar{x}_i-\bar{x})^2}{c-1}.$$

Foglaljuk össze eddigi eredményeinket: Van c db független, normális eloszlású, azonos szórású valószínűségi változónk, $\xi_1, \xi_2, \ldots, \xi_c$. Ezek várható értékére:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_c,$$

ill. a

$$\mu_i = \mu + t_i$$

modell felhasználásával

$$H_0: t_i = 0 \quad \text{minden } i\text{-re}, \quad i = 1, 2, \ldots, c.$$

A H_0 ellenőrzésére készítünk két független becslést a közös σ^2 varianciára. Ha H_0 igaz, akkor mind s^2, mind s_c^2 ugyanazon σ^2 becslése, és az $F = s_c^2/s^2$ hányados F-eloszlást követ, $c-1$, $n-c$ szabadsági fokpárral. Ha H_0 nem igaz, s^2 becslése lesz σ^2-nek, de s_c^2 értékét az egyes csoportok közötti szisztematikus eltérések (a kezelések hatása) növeli, így hányadosuk nagyobb lesz, mint a megfelelő szignifikanciaszinthez és $\{(c-1, n-c)\}$ szabadsági fokhoz tartozó F_{krit} érték. Ha $F > F^{\text{krit}}_{(\varepsilon, c-1, n-c)}$, akkor H_0-t nem fogadjuk el, így azt mondhatjuk, hogy az egyes csoportokban a várható értékek különbözőek, ezért a „kezelés" hatásosnak mondható ε szignifikanciaszinten.

A számolás megkönnyítésére és áttekinthetőbbé tételére szokás az ún. szórásfelbontó táblázat elkészítése. A módszer angol nevének (*AN*alysis *O*f *VA*riance) rövidítéseként elterjedt az ANOVA betűszó használata a szóráselemzés jelölésére.

Szóráselemzés egyszeres osztályozással
(Egyváltozós ANOVA táblázat)

Szóródás oka	Négyzetösszeg	Szabadsági fok	Szórásnégyzet
Csoportok között	$Q_1 = \sum_{i=1}^{c} n_i(\bar{x}_i - \bar{x})^2$	$c-1$	$s_c^2 = \frac{Q_1}{c-1}$
Csoportokon belüli (maradék)	$Q_2 = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	$n-c$	$s^2 = \frac{Q_2}{n-c}$
Teljes	$Q = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	$n-1$	

A szórásnégyzetnek az ANOVA táblázatban szereplő felbontása azzal a lényeges tulajdonsággal rendelkezik, hogy a teljes négyzetösszeg éppen a Q_1 és Q_2, azaz a csoportok közötti eltérés négyzetösszeg és a csoportokon belüli, ún. maradék négyzetösszeg összege, továbbá a teljes négyzetösszeg szabadsági foka éppen Q_1 és Q_2 szabadsági fokainak összege. (Ez a tulajdonság szükséges és elégséges feltétele annak, hogy s_c^2 és s^2 mindkettő független, χ^2 eloszlású valószínűségi változó legyen!)
Az állítás igazolására írjuk fel az egyes méréseknek az átlagtól való eltérését a következő módon:

$$x_{ij} - \bar{x} = (x_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{x}).$$

Emeljük négyzetre és összegezzük a kifejezés mindkét oldalát minden i, j-re:

$$\sum_{i,j} (x_{ij} - \bar{x})^2 = \sum_{i,j} (x_{ij} - \bar{x}_i)^2 + \sum_{i,j} (\bar{x}_i - \bar{x})^2 + 2 \sum_{i,j} (x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x}).$$

Vegyük észre, hogy a keresztszorzat zérus értékű, hiszen

$$\sum_{i=1}^{c}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)(\bar{x}_i-\bar{x}) = \sum_{i=1}^{c}\left\{(\bar{x}_i-\bar{x})\sum_{i=1}^{n_i}(x_{ij}-\bar{x}_i)\right\}$$

és $\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)=0$ minden i-re az átlag definíciója miatt. Ezt felhasználva éppen a kívánt azonossághoz jutunk.

A módszer illusztrálására vizsgáljuk meg a négy rugóacélgyár acélmintáira vonatkozó példát, és nézzük meg igaz-e az a hipotézis, hogy a minták szakítószilárdsága azonos, függetlenül attól, hogy melyik gyár gyártotta azokat. Mivel a szórásanalízis módszere alkalmazásának előfeltétele az, hogy az egyes csoportok szórása lényegében azonos legyen, hívjuk be először a STATISZTIKAI PRÓBÁK menüjéből az F-PRÓBA és BARTLETT-PRÓBA menüágat. Gépeljük be adatainkat. (A részleteket illetően l. a Bartlett-próbával foglalkozó 4.5.2 pontot.) A képernyőn a következő eredmény jelenik meg:

KHI-NÉGYZET =0,41565
SZABADSÁGI FOK=3
A VALÓSZÍNŰSÉG=0,063

A kapott eredmény szerint a szórások csak akkor volnának megkülönböztethetőek, ha (1−0,063)=93,7% kockázatot vállalnánk arra, hogy azonosak, de mi mégis (az eredmény véletlen jellege miatt) különbözőnek mondjuk őket. 5%-os kockázatvállalás (szignifikanciaszint) mellett a szórások azonosnak mondhatók, így a szóráselemzés elvégzésének nincs akadálya.

Térjünk át ezért a SZÓRÁSELEMZÉS menüágra és válasszunk ki a SZÓRÁSELEMZÉS EGYSZERES OSZTÁLYOZÁSSAL programot. Adatainkat ez esetben is akár lemezen előkészített file-ból, akár billentyűzetről bevihetjük. 4 mintacsoportunk van (4 kísérletsorozat, 4 „kezelés"), minden csoportban 10—10 mérési eredménnyel. Az adatok megadása után a képernyőn vagy kívánság esetén a nyomtatóval kiíratva a következő eredményt látjuk:

A SZÓRÁSFELBONTÓ TÁBLÁZAT

	NÉGYZET-ÖSSZEG	SZABADSÁGI FOK	SZÓRÁS-NÉGYZET
CSOPORTOK KÖZÖTTI	5151,875	3	1717,292
CSOPORTOKON BELÜLI	8347,500	36	231,875
TELJES	13 499,375	39	

F-ÉRTÉK = 7,40611
A CSOPORTOK KÖZÖTTI ELTÉRÉS 0,08%-os SZINTEN SZIGNIFIKÁNS

A megjelenő eredmény azt mutatja, hogy a kezelések várható értékei jelentősen eltérnek, már 0,1%-alatti szignifikanciaszinten is. (Ha ennél nagyobb kockázatot vállalunk arra, hogy a helyes nullhipotézist a véletlen miatt eltérő mintaátlagok (csoportátlagok) miatt elvetjük, akkor még inkább el kell vetnünk a várható értékek azonosságának feltételezését.) Az eredmény értelmezését megkönnyíti és a döntés jogosságát az Olvasóban megerősíti, ha a Melléklet 4. táblázatából kikeresi a (3,36) szabadsági fokú F elosztás kritikus értékét, mondjuk 1%-os szignifikanciaszinten

$$F_{(\varepsilon=0.01,\,3,\,36)} \cong 4{,}4.$$

Mivel a számított F-érték ennél nagyobb, a nullhipotézist el kell vetni.

SZÓRÁSELEMZÉS (EGYSZERES)

A változók listája:

L	— a mintacsoportok száma,
N(I)	— a mintacsoportok elemszáma,
X(J, I)	— a mintacsoportok elemei,
MM	— az összes elemek száma,
S1	— részeredmény,
S	— csoportok közötti négyzetösszeg,
SA	— S kerekítve,
S3	— csoportokon belüli négyzetösszeg,
SC	— S3 kerekítve,
S2	— négyzetösszeg összesen,
SE	— S2 kerekítve,
SB	— csoportok közötti korrigált tapasztalati szórásnégyzet,
SD	— csoportokon belüli korrigált tapasztalati szórásnégyzet,
SF	— f-érték,
L	— csoportok közötti szabadsági fok,
J	— csoportokon belüli szabadsági fok,
M, N	— lemezről beolvasott adatmátrix oszlopainak és sorainak száma,
OX	— mintacsoport adatmátrixbeli oszlopszáma,
A$	— Igen-Nem választás,
ID	— kiíratásnál választás (képernyő vagy nyomtató),
N$	— adatfile neve,
HS, HN$, S, SZ	— lemezhiba ellenőrzése.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—240:	A program neve.
250—1040:	Adatbevitel, számítások.
1050—1300:	Eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
1310—1520:	A program által nyújtott lehetőségek rövid ismertetése.
1530:	Visszatérés a FŐMENÜ c. programra.
1540—1960:	Adatbeolvasás lemezről. Közben lemezhiba ellenőrzése az 1970-es szubrutinnal.
1970—2080:	Lemezhibát ellenőrző szubrutin.
2090—2190:	Várakozó szubrutin.
2200—2270:	Várakozó szubrutin Igen-Nem válaszra.

```
900 REM*****************************
910 REM*                           *
920 REM*    SZAMITASOK             *
930 REM*                           *
940 REM*****************************
950 PRINT CHR$(147)
960 S1=0:S2=0:S3=0:MM=0:FOR I=1 TO L:S=0:FOR J=1 TO N(I):X=X(J,I)
970 S=S+X:S2=S2+X*X:NEXT J:S1=S1+S:S3=S3+S*S/N(I):MM=MM+N(I):NEXT I
980 S1=S1*S1/MM:S=S3-S1:S3=S2-S3:S2=S2-S1:J=MM-L:L=L-1
990 SA=INT(S*1E5)/1E5
1000 SB=INT(S/L*1E5)/1E5
1010 SC=INT(S3*1E5)/1E5
1020 SD=INT(S3/J*1E5)/1E5
1030 SE=INT(S2*1E5)/1E5
1040 SF=S/L*J/S3
```

5.2. Szóráselemzés kétszeres osztályozással

Tegyük fel, hogy egy adott növényfajta termeléséhez r-féle vetőmagot használhatunk fel. Szeretnénk tudni, hogy melyik vetőmag alkalmazása adja a legjobb terméshozamot. A termelőszövetkezet földjei azonban különböző minőségűek, más a talaj összetétele, pH-ja stb. Ez a terméshozamokat természetesen befolyásolhatja. Ezért ezt a hatást szeretnénk kiküszöbölni. Úgy is fogalmazhatnánk, hogy a vetőmagfajta hatását akarjuk tisztán értékelni, a talaj hatását pedig kiszűrni a vizsgálatból. Természetesen lehetne ez fordítva is, vagy akár kíváncsiak lehetnénk mindkét hatásra tisztán is. Mindez a módszer szempontjából nem játszik szerepet, de egyszerűbb a magyarázat, ha először egyetlen hatás (faktor) vizsgálatára szorítkozunk.

Ugyancsak az egyszerűség kedvéért tegyük fel először, hogy minden vetőmagfajtával, minden talajtípusban végzünk egy-egy kísérletet. Mérési adataink így táblázatos formába rendezhetők, minden vetőmagtípus és minden talajtípus kijelöli a táblázat egy elemét. A táblázat így kijelölt helyét *cellának* szokás nevezni. Azt mondhatjuk tehát, hogy kísérleteinket két faktor befolyásolja, adatainkat kétféle szempont szerint osztályozhatjuk, és cellánként egyetlen kísérletet végeztünk. Jelöljük az egyik hatást t_j-vel (az angol treatment (kezelés) szó kezdőbetűjét alkalmazva), a másik hatást b_i-vel (az angol block szóra emlékeztetve).

Ha feltételezzük, hogy a vetőmag fajtája és a talajtípus egymástól függetlenül hat a terméshozamra, akkor ezt a következő alakban fejezhetjük ki:

$$\mu_{ij} = \mu + t_j + b_i \quad (i = 1, 2, \ldots, r; \ j = 1, 2, \ldots, c),$$

ahol c a vetőmagfajták száma; r az alkalmazott talajtípusok száma.

Az előbbi felírással egyenértékű a mérési adatokra a következő modell lesz érvényes:

$$x_{ij} = \mu + b_i + t_j + \varepsilon_{ij} \quad (i = 1, \ldots, r;\ j = 1, \ldots, c).$$

Természetesen

$$\sum_{i=1}^{r} b_i = 0; \quad \sum_{j=1}^{c} t_j = 0.$$

Feltételezzük, hogy ε_{ij} a véletlen okozta hiba, amely $N(0, \sigma)$ eloszlású és ε_{ij}-k mind függetlenek. Hipotézisünk az, hogy a vetőmagfajtáknak (kezeléseknek) nincs hatása a várható terméshozamra, azaz

$$H_0: t_j = 0 \quad \text{minden } j\text{-re.}$$

Az egyszeres osztályozású szórásanalízishez hasonlóan itt is a teljes négyzetösszeg független négyzetösszegekre történő felbontását keressük. Az első négyzetösszeg a kezelések következtében fellépő változást méri, a második az egyes blokkok közötti eltérések négyzetösszege, a harmadik a maradék négyzetösszeg. A kezelések következtében fellépő szórást ezután a maradék szóráshoz viszonyítjuk, a szokásos F-próbát használva.

Vezessük be a következő jelöléseket:

$$\bar{x}_{i.} = \frac{1}{c}\sum_{j=1}^{c} x_{ij} \bar{x}_{.j} = \frac{1}{r}\sum_{i=1}^{r} x_{ij} \bar{x} = \frac{1}{rc}\sum_{j=1}^{c}\sum_{i=1}^{r} x_{ij}.$$

Ezekkel a jelölésekkel felírható a következő egyenlet:

$$Q = \sum_{i,j} (x_{ij} - \bar{x})^2 = c\sum_{i=1}^{r} (\bar{x}_{i.} - \bar{x})^2 + r\sum_{j=1}^{c} (x_{.j} - \bar{x})^2 +$$
$$+ \sum_{i,j} (x_{ij} - \bar{x}_{i.} - x_{.j} + \bar{x})^2 = Q_1 + Q_2 + Q_3.$$

Az egyszeres osztályozású szórásanalízisnél elmondott elvek alapján az előbbi azonosság bizonyítását az Olvasóra bízzuk. A kifejezés jobb oldalán álló első tag a blokkok közötti eltérés mértéke, a második a kezelések közöttié, a harmadik a maradék négyzetösszeg. Egyszerűen belátható az is, hogy az egyes négyzetösszegek szabadsági fokai a következők:

$$\begin{array}{l} Q_1:\ r-1 \\ Q_2:\ c-1 \\ Q_3:\ (r-1)(c-1) \\ \hline Q\ :\ rc-1 \end{array}$$

Osszuk most el rendre a négyzetösszegeket szabadsági fokaik számával!

$$s_1^2 = \frac{Q_1}{r-1} = \frac{c\sum_{i=1}^{r} (\bar{x}_{i.} - \bar{x})^2}{r-1},$$

$$s_2^2 = \frac{Q_2}{c-1} = \frac{r\sum_{j=1}^{c} (\bar{x}_{.j} - \bar{x})^2}{c-1},$$

$$s_3^2 = \frac{Q_3}{(r-1)(c-1)} = \frac{\sum\limits_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2}{rc-1}.$$

Belátható, hogy ezek várható értékei a következők:

$$M(s_1^2) = \sigma^2 + \frac{c}{r-1} \sum_{i=1}^{r} b_i^2,$$

$$M(s_2^2) = \sigma^2 + \frac{r}{c-1} \sum_{j=1}^{c} t_j^2,$$

$$M(s_3^2) = \sigma^2.$$

Sem a Q_2-t, sem Q_3-t nem befolyásolja, hogy a blokkok közötti eltérés négyzetösszege mekkora. Ha a nullhipotézis igaz, akkor minden t_j érték zérus, és $M(s_2^2) = M(s_3^2)$, így az $\left(\frac{s_2^2}{s_3^2}\right)$ hányados F-eloszlást követ $(c-1)$, $(r-1)$. $(c-1)$ szabadsági fokpárral, és így ez a hányados összehasonlítható adott szignifikanciaszinten az F-eloszlás kritikus értékeivel.

Ha $F_{(c-1),(r-1)(c-1)} < F_\varepsilon^{\text{krit}}$, akkor a hipotézis igaz. Ellenkező esetben szignifikáns különbség van az egyes vetőmagfajtákkal elérhető hozamok között, általánosan fogalmazva, az egyes „kezelések" várható értéke nem azonos. A kiszámítandó négyzetösszegeket és szabadsági fokaikat a számítás megkönnyítése érdekében táblázatosan is összefoglaltuk:

Szóráselemzés kétszeres osztályozással, cellánként 1 kísérlettel

A szóródás oka	Négyzetösszeg	Szabadsági fok	Becsült várható érték
Kezelés hatása (oszlopok eltérése)	$r \sum\limits_{j=1}^{c} (\bar{x}_{.j} - \bar{x})^2$	$c-1$	$\sigma^2 + \frac{r}{c-1} \sum\limits_j t_j^2$
Blokk hatás (sorok közötti eltérés)	$c \sum\limits_{i=1}^{r} (\bar{x}_{i.} - \bar{x})^2$	$r-1$	$\sigma^2 + \frac{c}{r-1} \sum\limits_i b_i^2$
Maradék	$\sum\limits_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} - \bar{x})^2$	$(r-1)(c-1)$	σ^2
Teljes	$\sum\limits_{i,j} (x_{ij} - \bar{x})^2$	$rc-1$	

Tegyük fel most, hogy cellánként nem egy, hanem n kísérletet végeztünk, azaz

$$x_{ijk} = \mu + b_i + t_j + \varepsilon_{ijk}$$

ahol: $i = 1, \ldots, r$; $j = 1, \ldots, c$; $k = 1, \ldots, n$.

Az előző gondolatmenetet követve levezethető, hogy a kiszámítandó négyzetösszegek és szabadsági fokaik a következők lesznek:

Szóráselemzés kétszeres osztályozással, cellánként n kísérlettel

A szóródás oka	Négyzetösszeg	Szabadsági fok	Becsült várható érték
Kezelések hatása (oszlopok)	$nr \sum_j (\bar{x}_{.j}-\bar{x})^2$	$c-1$	$\sigma^2+nr \sum_{j=1}^{c} \frac{t_j^2}{c-1}$
Blokkhatás (sorok)	$nc \sum_i (\bar{x}_{i.}-\bar{x})^2$	$r-1$	$\sigma^2+nc \sum_{i=1}^{r} \frac{b_i^2}{r-1}$
Maradék	különbségképzéssel	$nrc-c-r+1$	σ^2
Teljes	$\sum_{i,j,k} (x_{ijk}-\bar{x})^2$	$nrc-1$	

A vizsgált nullhipotézis és az F-próba alkalmazása a cellánkénti 1 kísérletnél elmondottakkal analóg. Ha H_0: $t_j=0$ minden j-re igaz, akkor a kezelésekre vonatkozó és a maradék négyzetösszegből kapott szórások hányadosa F-eloszlású, $(c-1)$, $(nrc-c-r+1)$ szabadsági fokpárral, és ha

$$F_{(c-1)(nrc-c-r+1)} < F_\varepsilon^{\text{krit}},$$

akkor a nullhipotézist elfogadjuk.

Az elmondottak megvilágítására tekintsük a következő példát: Három vetőmagfajtát kell összehasonlítanunk (V_1, V_2 és V_3) és 4 különféle típusú talajon kívánjuk ezeket kipróbálni. Mérési adatainkat (q/hektár) a következő táblázatban foglaltuk össze:

Talajtípus	Vetőmagtípus		
	V_1	V_2	V_3
1.	21	23	24
2.	18	17	23
3.	18	21	20
4.	17	20	22

A táblázatból látható, hogy cellánként egy kísérletet végeztünk. Nullhipotézisünk az, hogy a várható hozamok minden vetőmagfajtára azonosak. Válaszoljunk a kérdésre 5%-os szignifikanciaszinten, ill. határozzuk meg azt a szignifikanciaszintet, amelyen még éppen különbséget tudunk tenni a vetőmagtípusok szerint.

A feladat elvégzésére a STATISZTIKAI PRÓBÁK menüjéből válasszunk a SZÓRÁSELEMZÉS menüágat, abból pedig a SZÓRÁSELEMZÉS KÉTSZERES OSZTÁLYOZÁSSAL programágat. A választás hatására a program megkérdezi, hogy kérünk-e ismertetést. Ha kérünk, kiírja a képernyőre a módszer elvét, majd a mérési adatokat kéri. Ha nem, akkor egyenesen erre a pontra tér. Mérési adatainkat akár billentyűzetről, akár lemezen előkészített file-ból bevihetjük. A billentyűzetről való adatbevitelnél előbb az adatmátrix sorainak számát kéri (ez esetünkben az eltérő talajtípusok száma, azaz 4), majd a mátrix oszlopainak számát (a vetőmagtípusok számát), ami ez esetben 3. Az adatokat soronként kell megadni.

Az adatok megadása után a képernyőn a következő eredmény jelenik meg (ill. nyomtató választása esetén a nyomtatóval íratható ki).

A SZÓRÁSFELBONTÓ TÁBLÁZAT

	NÉGYZET-ÖSSZEG	SZABADSÁGI FOK	SZÓRÁS-NÉGYZET
Oszlop	28,17	2	14,08
Sor	22	3	7,33
Maradék	14,5	6	2,42
Teljes	64,67	11	
F (oszlop):	5,83		
F (sor):	3,03		

Elemezzük az előbbi eredményeket! Térjünk vissza a FŐMENÜ-höz, és válasszuk ki az eloszlás függvények menüjéből az F ELOSZLÁST! Adjuk meg először az oszlopok eltérésének F értékét (5,83), majd a számláló és nevező szabadsági fokainak számát (2 és 6). A képernyőre a következő eredmény íródik ki:

A SZIGNIFIKANCIASZINT 0,039

Az eredmény tehát azt jelenti, hogy az oszlopok közötti eltérés kb. 4%-os szignifikanciaszinten már szignifikáns. 5%-os szinten még inkább az. Így azt mondhatjuk, hogy a vetőmagfajtának szignifikáns hatása van a terméshozamra.

Adjuk meg most a sorok eltérésének F értékét (3,0) és szabadsági fokait (3,6). Ezt könnyűszerrel, a FŐMENÜ ismételt hívása nélkül is megtehetjük, mert a program megkérdezi a főmenühöz való visszatérés előtt, hogy akarunk-e még más értékkel is számolni. A képernyőn megjelenő eredményből megtudhatjuk, hogy az eltérő talajtípusok hatása csak 11,5%-os szignifikanciaszinten lenne kimutatható. 5%-os szignifikanciaszinten a nullhipotézist (azaz a talajtípusok hatásának lényegtelen voltát) el kell fogadnunk.

Megjegyzés: A programcsomag a kétszeres osztályozású szórásanalízist csak a cellánként 1 kísérletet megengedő esetre tartalmazza.

SZÓRÁSELEMZÉS KÉTSZERES OSZTÁLYOZÁSSAL

A változók listája:

N — az egyik faktor szintjeinek száma (a mátrix sorainak száma),
M — a másik faktor szintjeinek száma (a mátrix oszlopainak száma),
X(I, J) — az adatok,
T(J), S1, S2 — részeredmény,
S — teljes négyzetösszeg,
SA — S kerekítve,
S4 — oszlopok közötti négyzetösszeg,
SB — S4 kerekítve,
S3 — sorok közötti négyzetösszeg,
SD — S3 kerekítve,
S5 — eltérés négyzetösszege,
SG — S5 kerekítve,
V1 — oszlopok közötti korrigált tapasztalati szórásnégyzet,
SC — V1 kerekítve,
V2 — sorok közötti korrigált tapasztalati szórásnégyzet,
SE — V2 kerekítve,
V — eltérés korrigált tapasztalati szórásnégyzete,

SH	— V kerekítve,
I	— teljes szabadsági fok,
J	— oszlopok közötti szabadsági fok,
K	— sorok közötti szabadsági fok,
L	— eltérés szabadsági foka,
SF	— f-érték oszlopok között,
SI	— f-érték sorok között,
M, N	— lemezről beolvasott adatmátrix oszlopainak és sorainak száma,
R	— a vizsgált oszlopok száma,
O(I)	— a kijelölt oszlopok,
A$	— Igen-Nem választás,
ID	— kiíratásnál választás (képernyő vagy nyomtató),
N$	— adatfile neve,
HS, HN$, S, SZ	— lemezhiba ellenőrzése.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—240:	A program neve.
350—1040:	Adatbevitel, számítások.
1050—1340:	Eredmények kiírása választás szerint képernyőre vagy nyomtatóra.
1350—1520:	A program által nyújtott lehetőségek rövid ismertetése.
1530:	Visszatérés a FŐMENÜ c. programra.
1540—1950:	Adatbeolvasás lemezről. Közben lemezhiba ellenőrzése az 1960-as szubrutinnal.
1960—2070:	Lemezhibát ellenőrző szubrutin.
2080—2180:	Várakozó szubrutin.
2190—2260:	Várakozó szubrutin Igen-Nem válaszra.

```
840 REM*****************************
850 REM*                           *
860 REM*     SZAMITASOK            *
870 REM*                           *
880 REM*****************************
890 PRINT CHR$(147)
900 S1=0:S2=0:S3=0:S4=0:FOR I=1 TO N:S5=0:FOR J=1 TO R:X=X(I,J)
910 T(J)=T(J)+X:S5=S5+X:S2=S2+X*X:NEXT J
920 S1=S1+S5:S3=S3+S5*S5:NEXT I
930 FOR J=1 TO R:S4=S4+T(J)*T(J):NEXT J
940 S1=S1*S1/N/R:S=S2-S1:I=R*N-1:S4=S4/N-S1:J=R-1:V1=S4/J
950 S3=S3/R-S1:K=N-1:V2=S3/K:S5=S-S3-S4:L=I-J-K:V=1E37:IF L<>0 THEN V=S5/L
960 SA=INT(S*1E5)/1E5
970 SB=INT(S4*1E3)/1E3
980 SC=INT(V1*1E4)/1E4
990 SD=INT(S3*1E3)/1E3
1000 SE=INT(V2*1E4)/1E4
1010 SG=INT(S5*1E3)/1E3
1020 SH=INT(V*1E4)/1E4
1030 SI=1E37:IF V<>0 THEN SI=V2/V
1040 SF=1E37:IF V<>0 THEN SF=V1/V
```

6. REGRESSZIÓ ÉS KORRELÁCIÓSZÁMÍTÁS (GÖRBEILLESZTÉS)

A statisztikai adatfeldolgozásban gyakori feladat annak a megállapítása, hogy egy adott változó értéke függ egy másik változó (vagy más változók) értékétől. Az egyszerűség kedvéért tegyük föl először, hogy egy változó értékét egyetlen másik változó befolyásolja. Tehát 2 változó értékét mérjük egyidejűleg minden mérésnél. Az összetartozó mért értékeket azonos indexszel látjuk el. Ha n mérést végeztünk akkor n adatpárunk van: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, tömören fogalmazva $\{x_i y_i\}_{i=1}^n$.

Keressük azt a függvénykapcsolatot, amely kifejezi y függését x-től. Tipikus példa az ilyen vizsgálatokra a következő: Egy erdőgazdaságban fakitermeléssel foglalkoznak. Szeretnének egy egyszerű módszert találni a fakitermelés hozamának becslésére. Korábbi munkájukból sok mérési adatuk van: 100 esetben feljegyezték a kivágott fák magasságát és a fatörzs átmérőjét a vágás helyén. E két adatból a fa hasznos tömegét egyszerű képletekkel ki tudják számolni. A fák magasságát azonban csak a fa kivágása után lehet pontosan megmérni. Segítségünkre lehet az az ismert tény, hogy a telepített erdőkben a fák magassága arányos a fatörzs átmérőjével. Így kézenfekvő a gondolat, hogy célszerű lenne a kitermelhető fatömeget a könnyen mérhető fatörzsátmérő méretéből becsülni. Ehhez az szükséges, hogy a magasság becslését az átmérő mérésére vezessék vissza. Ez a visszavetítés, visszavezetés (idegen szóval regresszió) tehát egy olyan becslő függvény megkeresését jelenti, amellyel az egyik változó (a függő változó) értéke becsülhető a másik változó (a független változó) értékének ismeretében.

A regressziós módszerek tárgyalásánál az „alulról építkezés" módszerét követjük: egyszerűbb esetekben bemutatjuk a módszereket, majd bonyolultabb esetekre is kiterjesztjük a vizsgálatokat, végül bemutatjuk a módszerek hátterében megbúvó általános elméletet.

6.1. Egyváltozós lineáris regresszió a legkisebb négyzetek módszerével

Tegyük fel, hogy tudjuk, két változó között lineáris kapcsolat áll fenn. Tárgyalásunk elején azt is feltesszük, hogy az egyik változó értékét a kísérletező hiba nélkül állítja be, ill. tudja mérni, míg a másik, a függő változó értéke azonos független változó értéknél is kísérletről kísérletre ingadozik, azaz valamilyen érték körül ingadozva szóródást mutat. Más szavakkal, a függő változó valószínűségi változónak tekinthető.

Jelölje a független változót x, a függő változót y. A két változó közötti lineáris kapcsolat feltételezése azt jelenti, hogy fennáll az $y=a_0+a_1x$ összefüggés, ahol a_0 és a_1 a lineáris modell paraméterei. Az egyenlet egy egyenes egyenlete. Hogyan lehet az egyenes két paraméterét, a_0-t és a_1-et becsülni.

A becslés leggyakrabban használt, legelterjedtebb módszere a legkisebb négyzetek elvén alapul. Jelölje $\hat{a}_0$ az a_0 paraméter becsült értékét, $\hat{a}_1$ az a_1 paraméter becsült értékét, az összetartozó mérési adatokat pedig x_i és y_i. Az egyenes az x_i ponthoz az $\hat{a}_0+\hat{a}_1x_i$ értéket rendeli. Képezzük most a mért y_i és az x_i helyen számított $\hat{a}_0+\hat{a}_1x_i$ értékek különbségét:

$$e_i = y_i-(\hat{a}_0+\hat{a}_1x_i).$$

Az $\hat{a}_0$ és $\hat{a}_1$ értékét akkor határozzuk meg helyesen, ha ez a hiba, azaz a mért és számított értékek különbsége kicsi. Felmerül a kérdés, hogy az $e_1, \dots, e_i$ hibák milyen függvényét tekintsük a jó becslés kritériumának. A legkisebb négyzetek elve szerint az a_0 és a_1 értéket úgy válasszuk meg, hogy az e_i^2 értékek összegének lehető legkisebb értékre csökkentését tűzzük ki célul:

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i-(\hat{a}_0-\hat{a}_1x_i)\right)^2 \to \min_{\hat{a}_0,\, \hat{a}_1} .$$

A Q négyzetösszeg az $\hat{a}_0$ és $\hat{a}_1$ paraméterek függvénye, ugyanis x_i és y_i itt már konkrét számértékek. A Q $\hat{a}_0$ és $\hat{a}_1$ szerinti minimumát úgy kapjuk meg, hogy kiszámítjuk $\hat{a}_0$ és $\hat{a}_1$ szerinti parciális deriváltjait, azokat zérussal tesszük egyenlővé és megkeressük az így kapott egyenletek $\hat{a}_0$ és $\hat{a}_1$ gyökeit:

$$\frac{\partial Q}{\partial \hat{a}_0} = \Sigma 2(y_i-\hat{a}_0-\hat{a}_1x_i)(-1) = 0,$$

$$\frac{\partial Q}{\partial \hat{a}_1} = \Sigma 2(y_i-\hat{a}_0-\hat{a}_1x_i)(-x_i) = 0.$$

Átrendezve a következő ún. normálegyenlet-rendszerhez jutunk:

$$n\hat{a}_0+\hat{a}_1\Sigma x_i = \Sigma y_i,$$

$$a_0\Sigma x_i+\hat{a}_1\Sigma x_i^2 = \Sigma x_i y_i.$$

Az egyenletrendszer az $\hat{a}_0$ és $\hat{a}_1$ paraméterekben lineáris. Megoldása többféleképpen is lehetséges. Ha számítógéppel oldjuk meg az egyenletrendszert, akkor rendszerint a következő formulákat használjuk:

$$\hat{a}_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2},$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}.$$

Belátható, hogy az $\hat{a}_1$ fenti képletével egyenértékű az

$$\hat{a}_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2},$$

ill. az

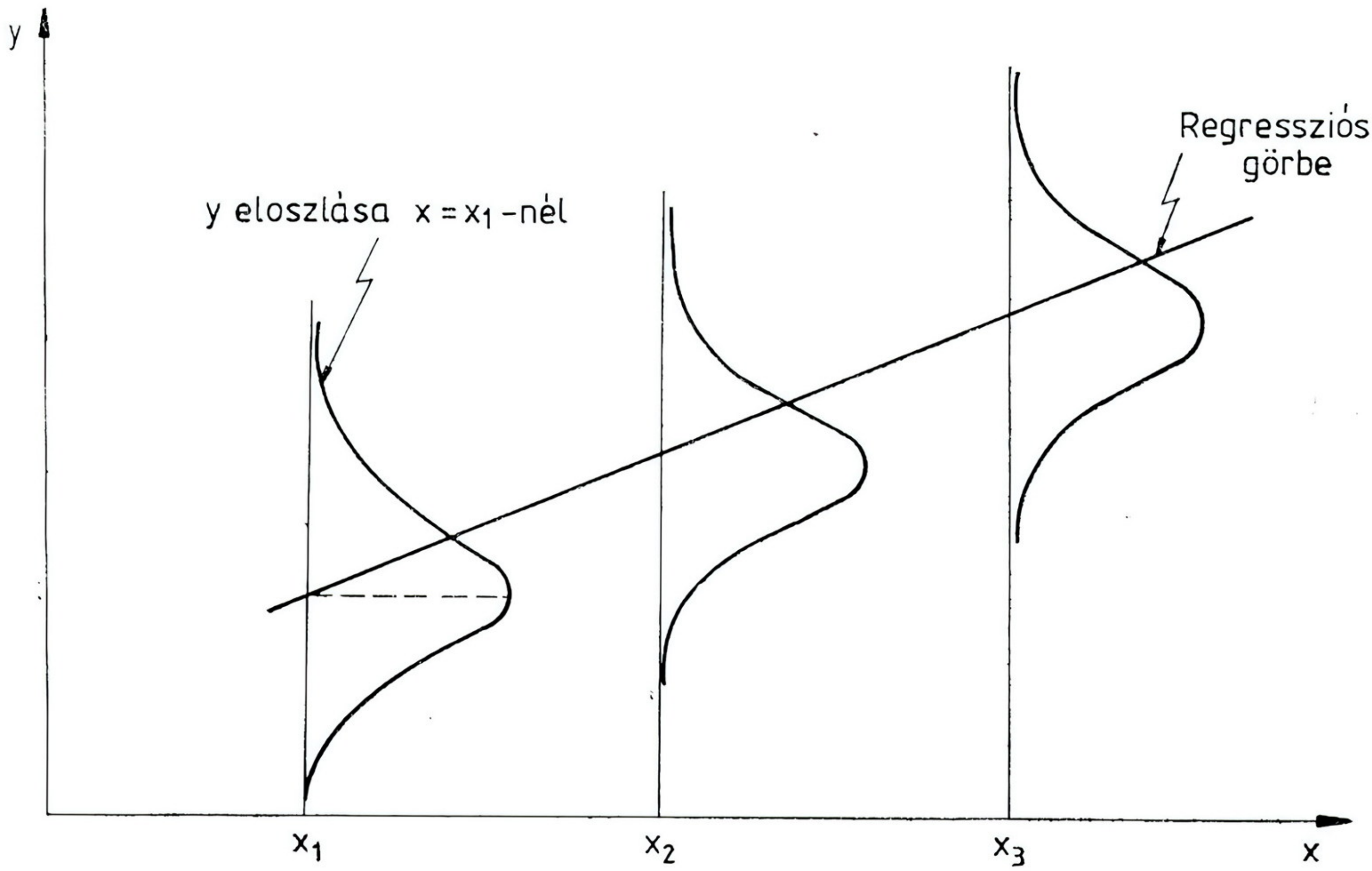
$$\hat{a}_1 = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2}$$

képlet is. Könnyen belátható az is, hogy a legkisebb négyzetek módszerével illesztett egyenes átmegy az $(\bar{x}, \bar{y})$ ponton, és meredeksége éppen $\hat{a}_1$. Ezért gyakran szokták a két ponton átmenő egyenes analitikus geometriából jól ismert képletét is használni az egyenes jellemzésére:

$$y - \bar{y} = \hat{a}_1(x - \bar{x}).$$

Ha $\hat{a}_0$ és $\hat{a}_1$ értékét meghatároztuk, tetszőleges x_0 értékhez kiszámíthatjuk az egyenesen fekvő y_0 pontot, azaz az y változó becslését az x_0 helyen feltéve, hogy x_0 a mérési pontok által lefedett intervallumba esett. (Az extrapoláció regressziós görbék esetében nem megengedhető.)

Ha minden x értéknél több mérést végeztünk, akkor a megfelelő y értékek minden x értéknél egy eloszlást adnak. Ezt a következő ábrával szemléltethetjük:



18. ábra. Regressziós görbe

Ha valóban az x és y változók közötti összefüggés egyenletét találtuk meg, és annak paramétereit határoztuk meg, akkor a regressziós görbe az y változó feltételes várható értékeit köti össze.

A regressziós feladat első problémája általában az egyenlet alakjának meghatározása, matematikai kifejezéssel élve a megfelelő függvényosztály kiválasztása. A végtelenül sokféle függvényosztály közül leggyakrabban az alábbiakat használjuk.

— A lineáris függvények osztálya: $y=a_0+a_1x$ egyetlen független változó esetén, $y=a_0+\sum_{i=1}^{k} a_i x_i$ k számú független változó esetén.

— A kvadratikus függvények osztálya: $y=a_0+a_1x+a_2x^2$ egyetlen független változóra.

— Az exponenciális függvények osztálya: $y=a_0e^x$ vagy $y=a_0x^{a_1}$ (egyváltozós eset).

A példákat tovább sorolhatnánk. A függvényalak megválasztását elméleti megfontolások befolyásolhatják, vagy a gyakorlatban igazolhatjuk egy-egy feltételezett összefüggés helyességét. A gyakorlatban úgy szokás eljárni, hogy többféle egyenletalakot is kipróbálunk, majd kiválasztjuk azt a függvényalakot, amely legjobban illeszkedik mérési pontjainkhoz.

Maradjunk egyelőre az egyváltozós, lineáris regresszió problémájánál. Milyen feltételeknek kell teljesülniük ahhoz, hogy a legkisebb négyzetek módszerével kapott $\hat{a}_0$ és $\hat{a}_1$ becsült paraméterek jó becslései legyenek a valódi összefüggés a_0 és a_1 paramétereinek?

Lineáris regressziónál kimutatható, hogy $\hat{a}_0$ és $\hat{a}_1$ a valódi paraméterek legnagyobb valószínűségű (ún. maximum likelihood) becslései, ha fennállnak a következő feltételek:

— a pontosan mérhető x változó minden konkrét értékénél az η valószínűségi változó normális eloszlású a_0+a_1x várható értékkel és $\sigma^2_{(\eta|x)}$ szórásnégyzettel,

— a $\sigma^2_{(\eta|x)}$ szórásnégyzet minden x értéknél azonos, azaz x értékétől független.

Ha a lineáris regressziós modellt ezzel a két feltétellel kiegészítjük, akkor a valódi regressziós egyenes, $y=a_0+a_1x$ éppen az

$$y = M(\eta|x) = a_0+a_1x$$

várható értékeket összekötő egyenes. A hibával terhelt mérési adatokból számított regressziós egyenest a valódi egyenestől való megkülönböztetés képpen

$$\hat{y} = \hat{a}_0+\hat{a}_1x\text{-szel jelöljük.}$$

Az előbbi két feltételt természetesen a módszer bevezetésében, a fejezet elején is leírhattuk volna (mint ahogy ezt a legtöbb szakkönyv is teszi). Sokéves tapasztalatunk szerint azonban ez azt a hamis látszatot keltené, hogy a legkisebb négyzetek módszerét egy egyenlet paramétereinek meghatározására csak akkor lehet használni, ha ezek a feltételek fennállnak. Ez azonban nem így van. Csupán annyit mondhatunk, hogy a paraméterbecslés akkor esik egybe az ún. maximum likelihood becsléssel (ami statisztikai értelemben egy jó, megbízható becslés) ha ezek a feltételek igazak. Ha például a második feltétel (a szórás függetlensége az x koordinátától) nem áll fenn, akkor a becslést azzal javíthatjuk, ha a kisebb szórású helyeken mért adatokat nagyobb súlyozó faktorral vesszük figyelembe az összegzésnél, ahol súlyozó faktorként a megfelelő x helyen vett szórás reciprokát vehetjük.

A regressziós egyenes statisztikai vizsgálata

A statisztikai vizsgálat első lépéseként nézzük meg, hogyan jellemezhető az egyenes illeszkedése a mérési pontokhoz! Az illeszkedés jóságának vizsgálatához bontsuk fel az y_i szórásának kifejezésében levő $\sum_{i=1}^{n}(y_i-\bar{y})^2$ ún. teljes négyzetösszeget két részre

$$Q = \sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}(y_i-\bar{y})^2 = Q_1+Q_2.$$

Az egyenletben $\hat{y}_i=\hat{a}_0+\hat{a}_1x_i$ a becsült paraméterekkel számított y_i érték az x_i helyen. (A felbontás jogosságát itt nem bizonyítjuk.)

Az egyenlet jobb oldalán

$$Q_1 = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

a mért és számított y értékek eltérésének négyzetösszege. Ez az a négyzetösszeg, amelyet minimalizálva az $\hat{a}_0$ és $\hat{a}_1$ értékeket meghatároztuk. A négyzetösszegnek ez már a minimuma, azaz azt mondhatjuk, hogy nem volt olyan $\hat{a}_0$ és $\hat{a}_1$ érték, amellyel ez még kisebbé tehető. Az eltérésnek ez a maradék része a regresszióval megmagyarázatlan maradt. Elnevezése ezért *maradék* négyzetösszeg. A jobb oldal másik tagja, $Q_2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ a számított $\hat{y}$ értékek és az $\bar{y}$ eltérése. Ez az a része a Q négyzetösszegnek, ami a regresszióra vezethető vissza, amit a regressziós egyenes magyaráz meg. (A regressziós egyenes is átmegy az $(\bar{x}, \bar{y})$ ponton, így akár $\hat{y}$-ot is írhattunk volna.) Q_2 neve *regressziós* négyzetösszeg.

Minél jobb az egyenes illeszkedése, annál kisebb Q_1, ill. annál nagyobb a Q_2/Q_1 hányados.

Bizonyítható, hogy Q eloszlása χ^2 eloszlás, szabadsági foka $n-1$. Mivel Q_1 szabadsági foka $n-2$, Q_2 szabadsági foka 1, azaz Q_1 és Q_2 szabadsági fokainak öszszege éppen Q szabadsági foka, ezért Q_1 és Q_2 független, χ^2 eloszlású valószínűségi változók.

Képezzük most az

$$F_{1,n-2} = \frac{\frac{Q_2}{1}}{\frac{Q_1}{n-2}}$$

hányadost. A hányados F-eloszlású, 1, $n-2$ szabadsági fokpárral. Ezért a hányados 1-től való eltérésére (a Q_2 és Q_1 szignifikáns eltérésére) statisztikai próbát konstruálhatunk. Nullhipotézisünk a következő:

$$H_0: Q_2 = Q_1.$$

Ha a nullhipotézis igaz, F-eloszlása $(1, n-2)$ szabadsági fokú Fisher-eloszlás. Kijelölve egy szignifikanciaszintet, és az F-próba kritikus értékével összehasonlítva eldönthető, hogy Q_2 és Q_1 eltérése szignifikáns-e. Ha Q_2 és Q_1 eltérése szignifikáns, azaz a nullhipotézist el kellett vetnünk, akkor az egyenlet illeszkedése megfelelőnek mondható (idegen szóval adekvát).

Másodszor feltehető az a kérdés, hogy a regresszió a teljes eltérés-négyzetösszeg

milyen hányadát magyarázza meg, azaz az összefüggés mennyire determinált. Az eltérés-négyzetösszeg, Q egy részét a regressziós egyenesen fekvő pontoknak az átlagtól való eltérésével sikerült megmagyarázni, ezt fejezi ki Q_2. Szokás ezért a Q_2/Q hányadosra a *determinációs koefficiens* elnevezést használni.

A determinációs koefficiens négyzetgyöke a *korrelációs együttható*. Ennek bővebb értelmezésére még visszatérünk.

Kimutatható, hogy a maradék négyzetösszeget, Q_1-et elosztva szabadsági fokával, éppen a $\sigma^2_{y/x}$ szórás torzítatlan becslését kapjuk.

$$S^2_{y|x} = \frac{(y_i - \hat{a}_0 - \hat{a}_1 x_i)^2}{n-2} = \frac{Q_1}{n-2}.$$

A fenti kifejezés négyzetgyökét a *becslés szórásának (standard deviációjának)* szokás nevezni a statisztikai iroldalomban.

A korrelációs együttható értelmezése

Térjünk vissza a fejezet elején vázolt alaphelyzethez: két változó értékét szimultán mérjük, így $\{x_i, y_i\}_{i=1}^n$ mérési adatsorunk van. A két változóról semmilyen előzetes feltevést nem teszünk, mindkettő valószínűségi változó. Legyen a két valószínűségi változó jele ξ és η. Arra a kérdésre keressük a választ, hogy van-e közöttük összefüggés. Kérdésünket egyelőre a lineáris összefüggés keresésére korlátozzuk.

A valószínűségszámításról szóló fejezetben beszéltünk a várható érték fogalmáról, ill. a valószínűségi változók függetlenségéről. Ha a ξ és η független valószínűségi változók, akkor a $[\xi - M(\xi)][\eta - M(\eta)]$ szorzat várható értéke 0-val egyenlő. Felmerül tehát az a gondolat, hogy tetszőleges ξ és η esetén ezek függőségét a $c = M\{[\xi - M(\xi)][\eta - M(\eta)]\}$ várható értékkel mérjük. Ezt a c számot a ξ és η valószínűségi változók *kovarianciájának* nevezzük. Attól függően, hogy milyen valószínűségi változókról van szó, a kovariancia $-\infty$ és $+\infty$ között bármely számértéket felvehet, a függőségnek viszont célszerű olyan mértéket választani, amely minden valószínűségi változó pár esetén abszolút határok közé esik. Bizonyítható, hogy a kovariancia abszolút értéke nem lehet nagyobb a $\sigma_\xi \sigma_\eta$ szorzatnál. A kovarianciát tehát ezzel elosztva mindig -1 és 1 közé eső számot kapunk. Jelöljük ezt a számot ϱ-val, akkor:

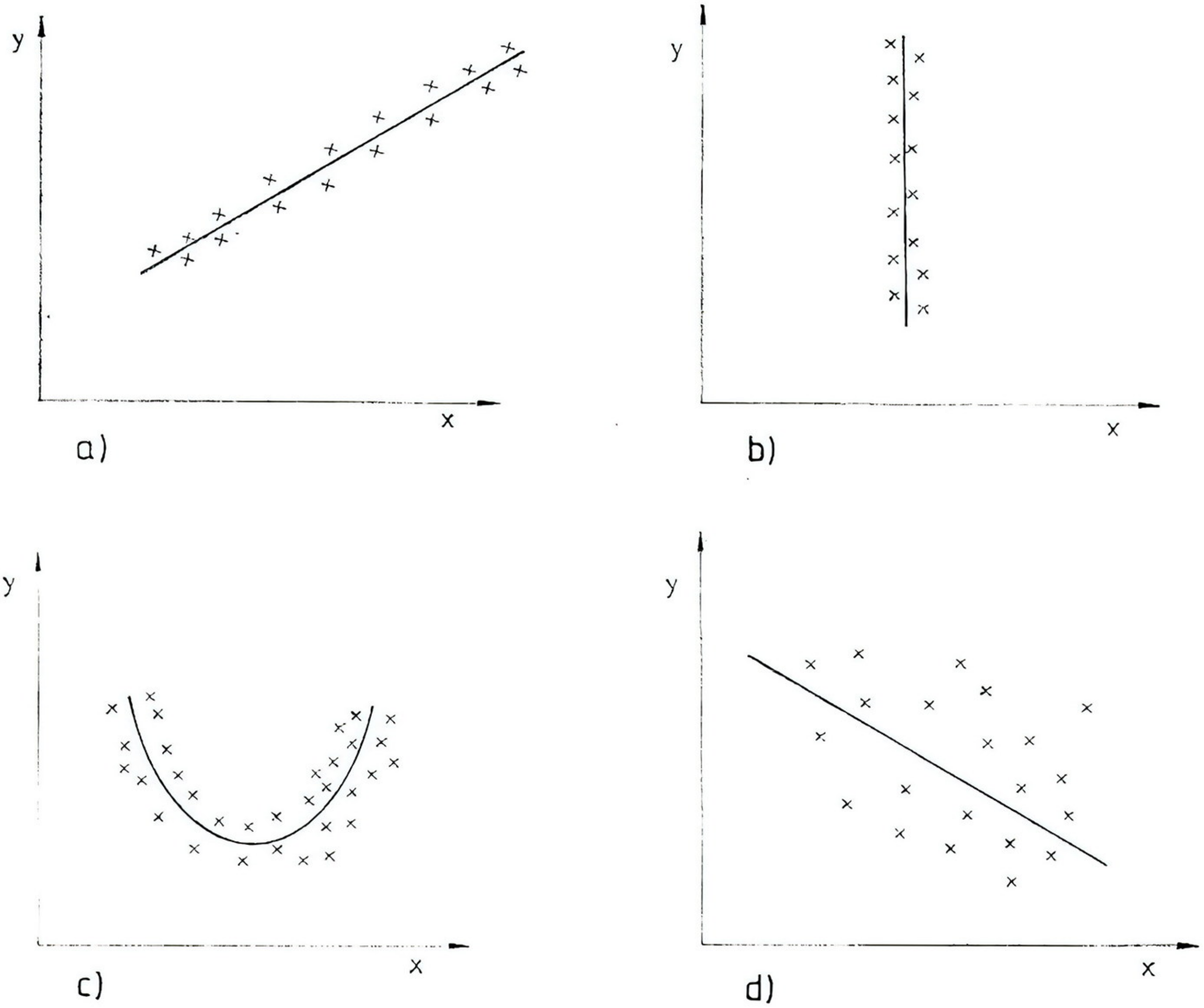
$$\varrho = \frac{M\{[\xi - M(\xi)][\eta - M(\eta)]\}}{\sigma_\xi \sigma_\eta}.$$

A ϱ neve: aξ és η valószínűségi változók korrelációs *együtthatója.*

Ha $\varrho = 0$, akkor ξ és η korrelálatlanok. A korrelálatlanság azonban nem jelent függetlenséget. Ez a korrelációs együttható egyik hibája. Másrészt $\varrho = 1$ akkor, és csak akkor, ha ξ és η között lineáris kapcsolat van, márpedig elegendő volna ξ és η között egy függvénykapcsolat fennállása is ahhoz, hogy a legnagyobb mértékű függőségről beszéljünk. Így tehát a korrelációs együttható a lineáris függőség mértéke, de nem mértéke általában a változók összefüggésének. A korrelációs együttható értéke és a változók közötti kapcsolat összefüggését a következő ábrákkal szemléltetjük (az egyszerűség kedvéért két változóra).

A korrelációs együtthatót a következő képlettel becsülhetjük:

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\Sigma(x_i - \bar{x})^2[\Sigma(y_i - \bar{y})]^2}},$$



19. ábra. A korrelációs együttható értéke és a változók közötti kapcsolat összefüggése
a) nagy pozitív korreláció; *b)* $\varrho \approx 0$; *c)* $\varrho \approx 0$; *d)* kis negatív korreláció

vagy az előbbi képlet azonos átalakításával, a szórások

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} \quad \text{és} \quad s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}}$$

jelölésének felhasználásával

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right).$$

Számítógépes számításoknál az

$$r = \frac{\Sigma x_i y_i - \dfrac{(\Sigma x_i)(\Sigma y_i)}{n}}{\sqrt{\left[\left(\Sigma x_i^2 - \dfrac{(\Sigma x_i^2)}{n}\right)\left(\Sigma y_i^2 - \dfrac{(\Sigma y_i)^2}{n}\right)\right]}}$$

képletet használjuk. A korrelációs együttható értéke annál szorosabb lineáris kapcsolatra utal, minél közelebb áll abszolút értéke 1-hez.

Megjegyzendő, hogy a korrelációs együttható képletében szerepel a kovariancia becslésére szolgáló képlet is. Jelöljük a kovariancát s_{xy}-nal:

$$s_{xy} = \frac{1}{n-1} \sum_{i-1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

Így a korrelációs együttható becsült értéke tulajdonképpen

$$r = \frac{s_{xy}}{s_x s_y}.$$

Felvethető az a kérdés, hogy az így kiszámított r korrelációs együttható mennyire megbízhatóan utal a két változó közötti (lineáris) összefüggésre. Erre az r abszolút értéke nem mindig elegendő információ, hiszen ha csak kevés mérésünk van, és nincs kapcsolat a két változó között, *véletlenül* lehetnek olyan adatpárjaink, amelyek viszonylag nagy (pozitív vagy negatív) korrelációt eredményeznek. Kérdés, hogy az így kapott r érték szignifikánsan különbözik-e a zérustól.

Ha a valódi korrelációs együttható zérus, és a mindkét változó normális eloszlású volt, akkor bizonyítható, hogy az

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

kifejezéssel definiált valószínűségi változó Student eloszlást követ $n-2$ szabadsági fokkal. Így a korrelációs együttható akkor különbözik szignifikánsan zérustól, ha értéke a megfelelő szignifikanciaszinten a t-próba kritikus értékénél nagyobb. Ha a nullától való eltérés szignifikanciája a kérdés, akkor természetesen elegendő az előbbi kifejezés abszolút értékét vizsgálni. Ez természetesen kétoldali próbához vezet, azaz α szignifikanciaszinten, a

$$H_0: |r| = 0$$

hipotézist akkor vetjük el, ha

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \geqq t_{(\varepsilon/2, n-2)}$$

Ha a szignifikanciaszintet a szokásos 5%-ban rögzítjük, akkor a mintanagyság függvényében $|r|$ kritikus értékeire a következőket kapjuk:

Mintanagyság	5	10	15	20	25	30	50	100
$|r|$ kritikus értéke	0,75	0,58	0,48	0,42	0,38	0,35	0,27	0,20

Ezzel minden szükséges előismeret birtokunkban van, ami egy egyszerű, egyváltozós lineáris regressziós példa számításához és eredményeinek értelmezéséhez szükséges. Lássunk tehát egy példát!

A különböző anyagok fajhője a hőmérséklet függvényében változik, általában a hőmérséklet növekedésével nő. Egy új típusú műanyag fajhőjének hőmérsékletfüggésére a következő adatokat mérték:

Hőmérséklet (°C)	0	10	20	30	40
Fajhő (J/(g · °C))	2,135	2,302	2,386	2,470	2,637

Mondhatjuk-e, hogy a fajhő a vizsgált hőmérséklet-tartományban lineárisan függ a hőmérséklettől. Illesszünk a mérési pontokra egyenest, ill. a 6 mérési pontra számítsuk ki az $y = A + Bx$ egyenes két paraméterét, A-t és B-t. Az egyenes paramétereinek ismeretében becsüljük a fajhő értékét a mérési pontokban!

A feladat elvégzéséhez a FŐMENÜ-ből az ADATFILE MŰVELETEK programot kiválasztva először vegyük fel adatainkat a FAJHŐ nevű file-ba. (A file-ba történő felvétel módjára itt nem térünk vissza, erre a 7. fejezetben látunk példát.) Ezután a REGRESSZIÓSZÁMÍTÁS menüjéből az EGYVÁLTOZÓS REGRESSZIÓ almenü LINEÁRIS programját választjuk a felkínált lehetőségek közül. Ezen a ponton kérhetünk ismertetést. Az ismertetés kérésére a következő szöveg jelenik meg a képernyőn:

A PROGRAM A LEGKISEBB NÉGYZETEK ELVE ALAPJÁN LINEÁRIS FÜGGVÉNYT ILLESZT A MÉRÉSI PONTOKRA. A FÜGGVÉNY ALAKJA: Y=A+B·X. AZ EGYENES ILLESZKEDÉSÉNEK JÓSÁGÁT IS VIZSGÁLJA, ÉS LEHETŐVÉ TESZI TETSZŐLEGES FÜGGETLEN VÁLTOZÓHOZ TARTOZÓ FÜGGVÉNYÉRTÉK KISZÁMÍTÁSÁT A BECSÜLT EGYENES SEGÍTSÉGÉVEL.

Ezután megkérdezi a program, hogy az adatok lemezen vannak-e (I/N). Természetesen lemezen vannak (I) és az ADAFILE NEVE kérdésre a FAJHŐ választ adjuk. Betettük az adatfile-t tartalmazó lemezt. A program ennek tudomásulvétele után megkérdezi a

FÜGGETLEN VÁLTOZÓ ADATMÁTRIXBELI OSZLOPSZÁMÁT (1)
és a FÜGGŐ VÁLTOZÓ ADATMÁTRIXBELI OSZLOPSZÁMÁT (2)

(Ezeket a kérdéseket azért teszi fel a program, hogy lehetőséget adjon más, a file-ban tárolt adatsorok közötti korrelációszámítására is. Például, esetünkben minden mérési ponthoz felvetettük volna a fajhőn kívül a sűrűség értékét is, és akkor egy másik regressziós számítással a sűrűség hőmérsékletfüggését vizsgálhatnánk.)

Az adatmátrix beolvasása után kiírathatjuk azt a képernyőre vagy a nyomtatóra, ellenőrizhetjük, módosíthatjuk az adatfile-kezelő program használatával stb.

A számítás eredményeit megjeleníthetjük a képernyőn vagy kiírathatjuk. Válasszuk most az előbbi megoldást. A képernyőn a következő kép jelenik meg:

A KÖZELÍTÉS Y=A+BX ALAKÚ
A=2,1516
B=0,0117
KÉRI A REGRESSZIÓS TÁBLÁZATOT? (I/N)

Írassuk ki a regressziós táblázatot is a képernyőre!

REGRESSZIÓS TÁBLÁZAT

REGRESSZIÓS NÉGYZETÖSSZEG	0,137
SZABADSÁGI FOK	1
NÉGYZET ÁTLAG	0,137
MARADÉK NÉGYZETÖSSZEG	$2{,}75 \cdot 10^{-3}$

SZABADSÁGI FOK 3
NÉGYZET ÁTLAG $9{,}185 \cdot 10^{-4}$
TELJES NÉGYZETÖSSZEG 0,14
SZABADSÁGI FOK 4
F=149,54 27 11
DETERMINÁCIÓS KOEFFICIENS = 0,980
KORRELÁCIÓS EGYÜTTHATÓ = 0,990
BECSLÉS SZÓRÁSA = 0,030
T (KORRELÁCIÓS EGYÜTTHATÓ)=12,2
NYOMJON MEG EGY BILLENTYŰT!

Mielőtt továbbmennénk, értékeljük az eredményeket! Az A és B paraméterek azt mutatják, hogy a fajhő értéke a vizsgált tartományban csak kismértékben változik a hőmérséklettel: 10°C hőmérsékletnövekedés kell ahhoz, hogy a fajhő 0,117 J/(g · °C)-szal, azaz kb. 5%-kal nőjön. Másrészt, ez a növekedés lineáris: a korrelációs együttható értéke 0,99 feletti, pedig az elméleti ismertetőből tudjuk, hogy 5 mérési pont esetén a korrelációs együttható már 0,75 felett szignifikáns 5%-os szignifikanciaszinten. Ez (természetesen) egybevág azzal az eredménnyel, hogy a regressziós együttható t-próba értéke 12,2, azaz jóval nagyobb a kritikus 5%-hoz tartozó 2,353 értéknél.

A regressziós táblázat teljes négyzetösszegének felbontása együttesen a determinációs koefficiensben, a becslés szórásában és az F-értékben jelenik meg.

A determinációs koefficiens 0,98-as értéke azt fejezi ki, hogy a fajhő változásának kb. 98%-át a hőmérséklet változására lehet visszavezetni, és a 2% a regressziós egyenletben figyelembe nem vett hatások eredménye. A becslés szórásának 0,03-os értéke (a maradék négyzetösszeg átlagértékének négyzetgyöke) azt mutatja, hogy a fajhő értékét 0,03-os abszolút pontossággal (szórással) tudjuk becsülni a kapott regressziós egyenessel.

Az F-próba 149,54-os értéke újabb bizonyítéka az egyenes igen jó illeszkedésének a mérési pontokhoz, hiszen a megfelelő kritikus érték

$$F_{(0,05,1,3)} = 10{,}13$$

és a kapott F-érték ennél egy nagyságrenddel nagyobb.

A programmal kiszámíthatjuk a fajhő értékét is különböző hőmérsékleteken. Példaképpen a mérési pontokban kiszámítva a fajhőket, értéküket összevethetjük a mért értékeikkel:

Hőmérséklet, °C	Fajhő, J/(g · °C)	
	Mért	Számított
0	2,135	2,1516
10	2,302	2,2688
20	2,386	2,386
30	2,470	2,5032
40	2,637	2,6204

Az egyváltozós lineáris regresszió könnyűszerrel továbbfejleszthető többféle irányban, többféle esetre is:

— ha a függő változó nemlineáris függvénye a független változónak, azaz az $Y=f(x)$ függvénykapcsolat x-ben nem lineáris;
— ha a függvénykapcsolat paramétereiben nem lineáris;
— ha a függő változó értékét több független változó is befolyásolja;
— ha a függő változó a független változó vagy változók valamilyen hatványaitól függ, de az összefüggés paramétereiben lineáris.

Az első két esetet (bizonyos feltételek fennállása esetén) visszavezethetjük az egyváltozós lineáris regresszióra, a második kettő a többváltozós lineáris regresszió ismeretét igényli, ezen belül a negyedik esetben a független változókat transzformálni kell, hogy a lineáris regresszió módszerét alkalmazni tudjuk.

6.2. Egyváltozós lineáris regresszió a független váltózó vagy a függvény transzformálásával

Abban az esetben, ha az illesztendő függvény paramétereiben nemlineáris, a lineáris regresszió módszere változatlan formában nem alkalmazható. Szigorúan véve az ún. nemlineáris regresszió módszeréhez kellene folyamodnunk. Ilyenkor a mért és számított értékek eltérés-négyzetösszegének a paraméterek szerinti minimumát általában nem lehet a parciális deriváltakat zérussal egyenlővé téve meghatározni. Ha azonban alkalmas transzformációval a függvény paramétereiben lineárissá tehető, akkor a linearizált függvény konstansainak becslésére alkalmazva a legkisebb négyzetek módszerét, a paraméterek meghatározására szolgáló egyenletrendszer, a normálegyenlet-rendszer lineáris lesz, így továbbra is a lineáris regresszió egyszerű eszköztárát használhatjuk. A konstansok meghatározása után a függvényt visszatranszformálva, az eredeti összefüggést kapjuk, a benne szereplő, most már ismert konstansokkal együtt. Figyelemmel kell természetesen lenni arra a tényre, hogy a különböző alkalmas transzformációk a mérési hibák eloszlását is transzformálják, így a becsült paraméterek eloszlása a legvalószínűbb érték körül kissé megváltozhat. Bár a transzformációk alkalmazása ezért nagy körültekintést igényel, mindez nem akadályozta meg az ilyen linearizáláson alapuló regressziós mószerek elterjedését. A következő táblázatban erre mutatunk be néhány példát.

Eredeti összefüggés	Transzformált		
	függvény	független változó	függő változó
$Y = A\exp(BX) \rightarrow$	$\ln Y = \ln A + BX$	$Y' = \ln Y$	$X' = X$
$Y = AX^B \rightarrow$	$\ln Y = \ln A + B\ln X$	$Y' = \ln Y$	$X' = \ln X$
$Y = A + B\log X \rightarrow$	$Y = A + B\log X$	$Y' = Y$	$X' = \log X$
$\ln Y = A + B/X \rightarrow$	$\ln Y = A + B/X$	$Y' = \ln Y$	$X' = 1/X$

A táblázat első sorában az exponenciális függvény lineáris regresszióval való illesztését látjuk. A független változót nem kellett transzformálni, a függő változót azonban igen. Ezzel a mérési hibák eloszlását is transzformáltuk. Az ilyen elven felépített program (EGYVÁLTOZÓS REGRESSZIÓ, EXPONENCIÁLIS) elsődlegesen az ln *A* és *B* értékét becsüli (természetesen nyomtatás előtt visszatranszformálva), és az erre a becslésre vonatkozó statisztikai táblázatot nyomtatja ki.

A táblázat második sorában olyan függvényt láthatunk, amelyben a becsülendő *B* paraméter a független változó kitevőjében van. Ennek lineáris regresszióval való illesztéséhez mind a független változót, mind a függő változót transzformálni kellett. A függő változó transzformálásával a mérési hibák eloszlását is transzformáltuk. Az ezt megvalósító program (EGYVÁLTOZÓS REGRESSZIÓ, GEOMETRIAI) elsődlegesen az ln *A* és a *B* értékét becsüli (természetesen nyomtatás előtt visszaformálva), és az erre a becslésre vonatkozó statisztikai táblázatot nyomtatja ki.

A táblázat harmadik sorában olyan függvényt láthatunk, amelyben a függő változó a független változó (tízes alapú) logaritmusától függ. Ennek lineáris regresszióval történő illesztéséhez csak a független változót kell transzformálni. A mérési hibák eloszlása így változatlan marad. Az ezt megvalósító program (EGYVÁLTOZÓS REGRESSZIÓ, LOGARITMIKUS) közvetlenül *A* és *B* értékét becsüli, a paramétereket így transzformálni nem kell.

A táblázat utolsó sorában egy ún. Antoine-egyenlet típusú függvény lineáris regresszióval történő illesztését látjuk. Ennek lineáris regresszióval való illesztéséhez mind a független változót, mind a függő változót transzformálni kellett. *A* függő változó transzformálásával mérési hibák eloszlását is transzformáltuk. Az ezt megvalósító program (EGYVÁLTOZÓS REGRESSZIÓ, ANTOINE) közvetlenül *A* és *B* értékét becsüli.

EGYVÁLTOZÓS REGRESSZIÓ
LINEÁRIS FÜGGVÉNY ILLESZTÉSE

A változók listája:

N	—	mérési pontok száma,
X(I)	—	független változó vektora,
Y(I)	—	függő változó vektora,
OX, OY	—	függő és független változó mátrixbeli oszlopszáma (lemezről történő adatbeolvasásnál használjuk),
A	—	0. fokú regressziós együttható,
B	—	1. fokú regressziós együttható,
S5	—	determinációs együttható,
S6	—	korrelációs együttható,
S7	—	a becslés standard hibája.

Programmagyarázat:

100:	Programcsomag-szervezés.
120—250:	A program neve.
260—720:	Adatbevitel.
730—860:	A regressziós függvény együtthatóinak meghatározása.
870—890:	Az eredmény kiírási helyének kijelölése.
900—1020:	Az együtthatók kiírása.

1030—1420:	A regressziós táblázat számítása.
1430—1580:	A mérési pontok ellenőrzése.
1590—1750:	Pontok meghatározása a regressziós görbéből.
1760:	Visszatérés a FŐMENÜ c. programra.
1770—2180:	Adatbeolvasás lemezről.
2190—2300:	Lemezhiba ellenőrzés.
2310—2460:	A program által nyújtott lehetőségek rövid ismertetése.
2470—2650:	Várakozás Igen-Nem válaszra.

```
730 REM******************************
740 REM*                             *
750 REM* EGYUTTHATOK MEGHATAROZASA  *
760 REM*                             *
770 REM******************************
780 FOR I=1 TO N
790 S1=S1+X(I)
800 S2=S2+Y(I)
810 S3=S3+X(I)↑2
820 S4=S4+Y(I)↑2
830 S5=S5+X(I)*Y(I)
840 NEXT I
850 B=(N*S5-S2*S1)/(N*S3-S1↑2)
860 A=(S2-B*S1)/N
```

EGYVÁLTOZÓS REGRESSZIÓ
EXPONENCIÁLIS FÜGGVÉNY ILLESZTÉSE

A változók listája:

N	— mérési pontok száma,
X(I)	— független változó vektora,
Y(I)	— függő változó vektora,
OX, OY	— függő és független változó mátrixbeli oszlopszáma (lemezről történő adatbeolvasásnál használjuk),
A	— preexponenciális tényező,
B	— a független változó együtthatója az illesztett függvényben ($Y = A * \exp(B * X)$),
S5	— determinációs együttható,
S6	— korrelációs együttható,
S7	— a becslés szórása.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—250:	A program neve.
260—720:	Adatbevitel.
730—860:	A regressziós függvény együtthatóinak meghatározása, a függő változó transzformációja $Y = \ln(Y)$.
870—890:	Az eredmény kiírási helyének kijelölése,
900—1020:	Az együtthatók kiírása.
1030—1420:	A regressziós táblázat számítása.
1430—1580:	A mérési pontok ellenőrzése.
1590—1750:	Pontok meghatározása a regressziós görbéből.
1760	Visszatérés a FŐNEMŰ programra.
1770—2180:	Adatbeolvasás lemezről.
2190—2300:	Lemezhiba-ellenőrzés.
2310—2460:	A program által nyújtott lehetőségek rövid ismertetése.
2470—2650:	Várakozás Igen-Nem válaszra.

```
730 REM*******************************
740 REM*                             *
750 REM* EGYUTTHATOK MEGHATAROZASA   *
760 REM*                             *
770 REM*******************************
780 FOR I=1 TO N:Y=LOG(Y(I))
790 S1=S1+X(I)
800 S2=S2+Y
810 S3=S3+X(I)*X(I)
820 S4=S4+Y*Y
830 S5=S5+X(I)*Y
840 NEXT I
850 B=(N*S5-S2*S1)/(N*S3-S1↑2)
860 A=EXP((S2-B*S1)/N)
```

EGYVÁLTOZÓS REGRESSZIÓ
$Y = Ax^B$ ALAKÚ FÜGGVÉNY ILLESZTÉSE

A változók listája:

N	—	mérési pontok száma,
X(I)	—	független változó vektora,
Y(I)	—	függő változó vektora,
OX, OY	—	függő és független változó mátrixbeli oszlopszáma (lemezről történő adatbeolvasásnál használjuk),
A	—	együttható ($Y = AX^B$),
B	—	kitevő,

S5 — determinációs együttható,
S6 — korrelációs együttható,
S7 — a becslés standard hibája.

Programmagyarázat:

100: Programcsomag-szervezés.
110—250: A program neve.
260—710: Adatbevitel.
730—860: A regressziós függvény együtthatóinak meghatározása, a független és függő változó transzformációja $X = \ln(X)$, $Y = \ln(Y)$.
870—890: Az eredmény kiírási helyének kijelölése.
900—1020: Az együtthatók kiírása.
1030—1420: A regressziós táblázat számítása.
1430—1580: A mérési pontok ellenőrzése.
1590—1750: Pontok meghatározása a regressziós görbéből.
1760: Visszatérés a FŐMENÜ c. programra.
1770—2180: Adatbeolvasás lemezről.
2190—2300: Lemezhiba-ellenőrzés.
2310—2460: A program által nyújtott lehetőségek rövid ismertetése.
2470—2650: Várakozás Igen-Nem válaszra.

```
730 REM********************************
740 REM*                              *
750 REM* EGYUTTHATOK MEGHATAROZASA    *
760 REM*                              *
770 REM********************************
780 FOR I=1 TO N:X=LOG(X(I)):Y=LOG(Y(I))
790 S1=S1+X
800 S2=S2+Y
810 S3=S3+X*X
820 S4=S4+Y*Y
830 S5=S5+X*Y
840 NEXT I
850 B=(N*S5-S2*S1)/(N*S3-S1↑2)
860 A=EXP((S2-B*S1)/N)
```

EGYVÁLTOZÓS REGRESSZIÓ LOGARITMUSFÜGGVÉNY ILLESZTÉSE

A változók listája:

N — mérési pontok száma,
X(I) — független változó vektora,
Y(I) — függő változó vektora,
OX, OY — függő és független változó mátrixbeli oszlopszáma (lemezről történő adatbeolvasásnál használjuk),

A	—	az illesztett függvény $(Y = A + B * \lg(X))$, konstans tagja,
B	—	a logaritmusfüggvény szorzója,
S5	—	determinációs együttható,
S6	—	korrelációs együttható,
S7	—	a becslés standard hibája.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—250:	A program neve.
330—720:	Adatbevitel.
730—860:	A regressziós függvény együtthatóinak meghatározása, a független változó transzformációja $X = \log(X)$.
870—890:	Az eredmény kiírási helyének kijelölése.
900—1020:	Az együtthatók kiírása.
1030—1420:	A regressziós táblázat számítása.
1430—1580:	A mérési pontok ellenőrzése.
1590—1750:	Pontok meghatározása a regressziós görbéből.
1760:	Visszatérés a FŐMENÜ c. programhoz.
1770—2180:	Adatbeolvasás lemezről.
2190—2300:	Lemezhiba-ellenőrzés.
2310—2460:	A program által nyújtott lehetőségek rövid ismertetése.
2470—2650:	Várakozás Igen-Nem válaszra.

```
730 REM*******************************
740 REM*                             *
750 REM* EGYÜTTHATOK MEGHATAROZASA   *
760 REM*                             *
770 REM*******************************
780 FOR I=1 TO N:X=LOG(X(I))/LOG(10)
790 S1=S1+X
800 S2=S2+Y(I)
810 S3=S3+X*X
820 S4=S4+Y(I)*Y(I)
830 S5=S5+X*Y(I)
840 NEXT I
850 B=(N*S5-S2*S1)/(N*S3-S1↑2)
860 A=(S2-B*S1)/N
```

EGYVÁLTOZÓS REGRESSZIÓ
ANTOINE TÍPUSÚ FÜGGVÉNY ILLESZTÉSE

A változók listája:

N	—	mérési pontok száma,
X(I)	—	független változó vektora,
Y(I)	—	függő változó vektora,
OX, OY	—	függő és független változók mátrixbeli oszlopszáma (lemezről történő adatbeolvasásnál használjuk),
A	—	az illesztett $(\ln(Y)=A+B*1/X)$ függvény konstans tagja,
B	—	a független változó reciprokának együtthatója,
S5	—	determinációs együttható,
S6	—	korrelációs együttható,
S7	—	a becslés standard hibája.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—250:	A program neve.
330—720:	Adatbevitel.
730—860:	A regressziós függvény együtthatóinak meghatározása, a független és függő változó transzformációja $X=1/X$, $Y=\ln(Y)$.
870—890:	Az eredmény kiírási helyének kijelölése.
900—1020:	Az együtthatók kiírása.
1030—1420:	A regressziós táblázat számítása.
1430—1580:	A mérési pontok ellenőrzése.
1590—1750:	Pontok meghatározása a regressziós görbéből.
1760:	Visszatérés a FŐMENÜ c. programhoz.
1770—2180:	Adatbeolvasás lemezről.
2190—2300:	Lemezhiba-ellenőrzés.
2310—2460:	A program által nyújtott lehetőségek rövid ismertetése.
2470—2650:	Várakozás Igen-Nem válaszra.

```
730 REM********************************
740 REM*                              *
750 REM* EGYUTTHATOK MEGHATAROZASA    *
760 REM*                              *
770 REM********************************
780 FOR I=1 TO N:X=1/X(I):Y=LOG(Y(I))
790 S1=S1+X
800 S2=S2+Y
810 S3=S3+X*X
820 S4=S4+Y*Y
830 S5=S5+X*Y
840 NEXT I
850 B=(N*S5-S2*S1)/(N*S3-S1↑2)
860 A=(S2-B*S1)/N
```

6.3. Többváltozós lineáris regresszió

Terjesszük ki egyváltozós lineáris modellünket oly módon, hogy a függő változó ne egyetlen független változótól függjön, hanem több független változó lineáris kombinációjától.

Példaképpen legyen 3 független, pontosan mérhető vagy pontosan ismert értékre beállítható független változónk, X_1, X_2, X_3. Az összefüggés pedig legyen lineáris:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3.$$

Annak érdekében, hogy a_i $(i=1, 2, 3)$ legnagyobb valószínűségű becslését kapjuk, szokás feltételezni, hogy az Y normális eloszlású és szórása az X_1, X_2, X_3 koordinátáktól független. (Ahogy az előzőekben is ezt a feltételezést tettük az egyváltozós esetre.)

Ha N mérést végeztünk, azaz van egy $[Y_i, X_{1i}, X_{2i}, X_{3i}]_{i=1}^{N}$ mérési adathalmazunk, ahol $N > 4$, az a_0, a_1, a_2, a_3 becslését az előző fejezetben ismertetett módon kaphatjuk. Felírjuk a

$$Q = \sum_{i=1}^{N} (Y_i - \hat{a}_0 - \hat{a}_1 x_{1i} - \hat{a}_2 x_{2i} - \hat{a}_3 x_{3i})^2$$

négyzetösszeget, és képezzük az $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$ szerinti parciális deriváltakat. Ezeket nullával egyenlővé téve szimultán (az $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$ paraméterekben) lineáris egyenletet kapunk:

$$\hat{a}_0 N + \hat{a}_1 \Sigma x_{1i} + \hat{a}_2 \Sigma x_{2i} + \hat{a}_3 \Sigma x_{3i} = \Sigma y_i,$$

$$\hat{a}_0 \Sigma x_{1i} + \hat{a}_1 \Sigma x_{1i}^2 + \hat{a}_2 \Sigma x_{2i} x_{1i} + \hat{a}_3 \Sigma x_{3i} x_{1i} = \Sigma y_i x_{1i},$$

$$\hat{a}_0 \Sigma x_{2i} + \hat{a}_1 \Sigma x_{1i} x_{2i} + \hat{a}_2 \Sigma x_{2i}^2 + \hat{a}_3 \Sigma x_{3i} x_{2i} = \Sigma y_i x_{2i},$$

$$\hat{a}_0 \Sigma x_{3i} + \hat{a}_1 \Sigma x_{1i} x_{3i} + \hat{a}_2 \Sigma x_{2i} x_{3i} + \hat{a}_3 \Sigma x_{3i}^2 = \Sigma y_i x_{3i}.$$

Ennek a normálegyenlet-rendszernek a megoldása adja az $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$ paraméterek becslését. Az egyenletrendszer megoldása tetszőleges (lineáris egyenletrendszer-megoldó) módszerrel végezhető. Programunk az ún. Gauss eliminációt alkalmazza. A regressziós egyenlet statisztikai vizsgálata hasonló módszerekkel lehetséges, mint egyváltozós esetben: a Q négyzetösszeget ugyanúgy bontjuk $Q1$ és $Q2$-re, mint egyváltozós esetben, csupán a négyzetösszegek szabadsági foka változik a paraméterek számának változása miatt: ha a független változók száma k:

— $Q1$ (a maradék négyzetösszeg) szabadsági foka $N-k-1$;

— $Q2$ (a regressziós négyzetösszeg) szabadsági foka k.

Ezt figyelembe véve, az egyenlet illeszkedését mérni hivatott F-próba hasonló módon végezhető el, mint egyváltozós esetben. Mivel a determinációs együttható, a korrelációs együttható, a becslés standard deviációja és t-próbája az előbbi négyzetösszegek alapján megkapható, ezek ismételt felírására nincs szükség. A determinációs együttható ez esetben természetesen azt fejezi ki, hogy az összes független változónak a regressziós egyenletben tükröződő változása a teljes négyzetösszegnek milyen hányadát képes megmagyarázni. A korrelációs együttható is értelemszerűen többváltozós korrelációs együtthatóvá válik. A becslés standard hibájának képletében a nevezőbe természetesen $Q1$ szabadsági foka, $N-k-1$ kerül, és ennyi lesz a megfelelő t-próba szabadsági fokainak száma is.

Lássunk most az elmondottak illusztrálására ismét egy példát!

Egy salétromsavüzemben ammónia oxidációjával állítanak elő salétromsavat. A nitrogén-oxidokat híg salétromsavval nyeletik el. Az abszorpció nem tökéletes: a véggázokkal nitrogén-oxidok távoznak. Passzív kísérletekben mérték az oxidáló levegőáram mennyiségét (x_1), a hűtővíz hőmérsékletét (x_2), az abszorbersav koncentrációját (amelyet a könnyebb kódolás kedvéért úgy kódoltak, hogy 50-et levontak belőle és 10-zel megszorozták) (x_3), és a nitrogénveszteségből számított százalékos ammóniaveszteséget. A függőváltozó (Y) ez utóbbi mennyiség tízszerese. 21 mérés adatait a következő táblázat tartalmazza.

Sorszám	x_1	x_2	x_3	Y
1	80	27	89	42
2	80	27	88	37
3	75	25	90	37
4	62	24	87	28
5	62	22	87	18
6	62	23	87	18
7	62	24	93	19
8	62	24	93	20
9	58	23	87	15
10	58	18	80	14
11	58	18	89	14
12	58	17	88	13
13	58	18	82	11
14	58	19	93	12
15	50	18	89	8
16	50	18	86	7
17	50	19	72	8
18	50	19	79	8
19	50	20	80	9
20	56	20	82	15
21	70	20	91	15

Feltételezve, hogy az ammóniaveszteség lineárisan függ a figyelembe vett független változóktól, írjuk le Y változását x_1, x_2 és x_3 függvényében.

A feladatot a REGRESSZIÓSZÁMÍTÁS programág TÖBBVÁLTOZÓS LINEÁRIS, A VÁLTOZÓK TRANSZFORMÁCIÓJA NÉLKÜL programjával oldhatjuk meg.

Tegyük fel, hogy adatainkat előzőleg az adatkezelő programmal egy AMMON nevű file-ba tettük. Ekkor a program bejelentkezik, és megkérdezi, kérünk-e ismertetést. Ha nem kértünk, vagy már megkaptuk és továbblapoztunk, akkor megkérdezi, hogy adatainkat lemezről vagy billentyűzetről akarjuk-e bevinni. Esetünkben lemezről. Ekkor megkér, hogy tegyük be az adatlemezt, majd ennek nyugtázása után beolvassa az adatokat, és megkérdezi, hogy kiírja-e az adatmátrixot. Ha ezt nem kérjük vagy kérjük, de már kiírtuk az adatokat a képernyőre (vagy nyomtatóra), és azokat megfelelőnek ítélve tovább folytattuk a program végrehajtását, akkor a program elvégzi a szükséges számításokat. Megkérdezi, hova kérjük az eredményeket, majd kiírja a választott perifériára előbb a kiszámított paramétereket, majd a regressziós egyenlet statisztikai vizsgálatának eredményét, esetünkben a következőket:

$b(0) = -39{,}9196$
$b(1) = \ \ 0{,}71564$
$b(2) = \ \ 1{,}29529$
$b(3) = -0{,}15212$

REGRESSZIÓS	
NÉGYZETÖSSZEG	1890,408
SZABADSÁGI FOK	3
NÉGYZET ÁTLAG	630,1
MARADÉK	
NÉGYZETÖSSZEG	178,8
SZABADSÁGI FOK	17
NÉGYZET ÁTLAG	10,5
TELJES	
NÉGYZETÖSSZEG	2069,2
SZABADSÁGI FOK	20

$F = 59{,}9022$

DETERMINÁCIÓS KOEFFICIENS = 0,914
TÖBBSZÖRÖS KORR. KOEFFICIENS = 0,956
BECSLÉS STANDARD DEVIÁCIÓJA = 3,243

KISZÁMÍTSUK *Y* ÉRTÉKÉT KÜLÖNBÖZŐ *X* ÉRTÉKEKRE?

Ez utóbbi kérdés megválaszolását az olvasó szabad belátására bízzuk. A regressziós egyenlet statisztikai vizsgálatából látszik, hogy

— F értéke jóval nagyobb a szokásos 5%-os szignifikanciaszintű kritikus értéknél $F(p, k, n-k-1) = F(0{,}05, 3{,}17) = 3{,}2$-nél;
— a determinációs koefficiens szerint a változás több, mint 91%-át leírja a modell;
— a korrelációs koefficiens elegendően nagy;
— a becslés standard deviációja nem túl nagy.

TÖBBVÁLTOZÓS LINEÁRIS REGRESSZIÓ

A változók listája:

N	— mérési pontok száma,
M	— független változók száma,
X(I, J)	— függő és független változók mátrixa ($I=1$-től N-ig, $J=1$-től $M+1$-ig megy). Az $M+1$. oszlopban van a független változó,
Q(I, J)	— az egyenlet együtthatói ($Y=A(1)*X1+A(2)*X2+\ldots+A(M)*XM$),
S5	— determinációs együttható,
S6	— korrelációs együttható,
S7	— a becslés standard hibája,
N$	— az adatfile neve.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—230:	A program neve.
240—780:	Adatbevitel.
810—1030:	A regressziós függvény együtthatóinak meghatározása.
1040—1060:	Az eredmény kiírási helyének kijelölése.
1070—1210:	Az együtthatók kiírása.
1220—1600:	A regressziós táblázat számítása.

1610—1780:	A mérési pontok ellenőrzése.
1790—1960:	Pontok meghatározása a regressziós görbéből.
1970:	Visszatérés a FŐMENÜ c. programhoz.
1980—2300:	Adatbeolvasás lemezről.
2310—2420:	Lemezhiba-ellenőrzés.
2430—2640:	A program által nyújtott lehetőségek rövid ismertetése.
2650—2830:	Várakozás Igen-Nem válaszra.

```
810 REM********************************
820 REM*                              *
830 REM* EGYUTTHATOK MEGHATAROZÁSA    *
840 REM*                              *
850 REM********************************
860 HS=0
870 FOR I=1 TO N:X(I,0)=1:NEXT I:X(0,M+1)=1
880 FOR I=1 TO N:FOR J=0 TO M:FOR K=0 TO M
890 Q(J,K)=Q(J,K)+X(I,J)*X(I,K):NEXT K:Q(J,M+1)=Q(J,M+1)+X(I,J)*X(I,M+1):NEXT J
900 D(M+1)=D(M+1)+X(I,M+1)*X(I,M+1):NEXT I
910 FOR I=0 TO M:D(I)=Q(I,M+1):E(I)=Q(0,I):NEXT I
920 FOR S=0 TO M:IF Q(S,S)<>0 THEN 980
930 IF S=M THEN 970
940 FOR T=S+1 TO M:IF Q(T,S)=0 THEN 960
950 FOR J=S TO M+1:B=Q(S,J):Q(S,J)=Q(T,J):Q(T,J)=B:NEXT J:T=S+1
960 NEXT T
970 IF Q(S,S)=0 THEN PRINT "NINCS EGYERTELMU MEGOLDAS!":HS=1:S=M:GOTO 1020
980 C=1/Q(S,S):FOR J=S+1 TO M+1:Q(S,J)=C*Q(S,J):NEXT J
990 FOR T=0 TO M:IF T=S THEN 1010
1000 C=-Q(T,S):FOR J=S+1 TO M+1:Q(T,J)=Q(T,J)+C*Q(S,J):NEXT J
1010 NEXT T
1020 NEXT S
1030 IF HS<>0 THEN 1940
```

6.4. Többváltozós lineáris regresszió a változók transzformációjával

Közelítés egyetlen változó hatványsorával (egyváltozós polinomiális regresszió)

Az előzőekben először az egy- (független) változós, majd a több- (független) változós lineáris regresszióval foglalkoztunk. Most ismét visszatérünk az egy- (független) változós esetre, de most ennek (a független változóban) nemlineáris esetét visszavezetjük a többváltozós lineáris regresszió esetére.

Tegyük fel, hogy a függő változó az (egyetlen) független változó valamilyen polinomjától függ. Mivel a legtöbb függvény hatványsorral jól közelíthető, ez a

feltételezés egyváltozós függvényeknél általában célravezető. Harmadfokú polinommal való közelítés esetén például a regressziós függvény alakja a következő lesz:

$$Y = a_0 + a_1 X + a_2 X^2 + a_3 X^3.$$

Az ilyen típusú regresszió teljesen hasonló módon kezelhető, mint a többváltozós lineáris eset, csupán a független változók transzformációjára van szükség. Legyen $x1 = X$, $x2 = X^2$, $x3 = X^3$. Az a_0, a_1, a_2, a_3 legkisebb négyzetek módszerével történő becslését ekkor a 6.3. alfejezetben szereplő többváltozós lineáris regressziós normálegyenletekből kapjuk.

Lineáris regresszió a független változók tetszőleges, de paramétereikben lineáris kifejezéseinek lineáris kombinációjával

Ugyanarra a mérési adatsorra különböző regressziós modelleket kipróbálva, eljuthatunk oda, hogy nemcsak több változót kell figyelembe vennünk, de ezek szorzatai, pozitív és negatív kitevős hatványai, sőt ezek egyéb kifejezései is bekerülnek a modellbe. Ez nem meglepő, hiszen a többváltozós függvények sorbafejtésénél is ilyen kifejezések írják le a függvény hatványsorát. Ez egyszerűbb esetben a többváltozós lineáris és az egyváltozós polinomiális regresszió kombinációjaként fogható fel. Tekintsük például a következő modelleket

$$Y = a_0 + a_1 X + a_2 X^2 + a_3 Z,$$

$$Y = a_0 + a_1 X + a_2 Z + a_3 XZ,$$

$$Y = a_0 + a_1 (1/X) + a_2 \log X + a_3 Z.$$

Az első esetben a többváltozós lineáris regressziót $X_1 = X$, $X_2 = X^2$, $X_3 = Z$ helyettesítéssel használhatjuk. A második esetben a megfelelő helyettesítések felírását az Olvasóra bízzuk. Megjegyezzük, hogy az XZ kifejezés két független változó kölcsönhatását fejezi ki. A harmadik esetben $X_1 = 1/X$, $X_2 = \log X$, $X_3 = Z$ helyettesítés vezet célra. Bár $\log X$ az X-nek nem hatványa, de ez a módszer alkalmazását nem befolyásolja, hiszen ezt a tényt sehol sem használtuk ki. Mivel ez az általános eset igen gyakran előfordul, erre külön programot írtunk. A program még arra is lehetőséget ad, hogy a függő változót is transzformáljuk. (Ilyenkor természetesen tudatában kell lenni annak, hogy a mérési hibák eloszlását is transzformáljuk!)

Nézzük a program alkalmazását a 6.3. alfejezetben bemutatott AMMON adatsor példáján! Tegyük fel, hogy a legjobban kilógó 21. pont elhagyása mellett döntöttünk (ennek jogosságáról az Olvasó a mért és az első modellel számított eredmények összevetésével meggyőződhet), és a többváltozós kovarianciaanalízisből (l. a 6.5. alfejezetben) azt a következtetést vontuk le, hogy az x_3 változó hatása az Y-ra nem szignifikáns. Helyette az x_2 négyzetével próbálkozunk. Ha ez a csere a többváltozós korrelációs együtthatót, a determinációs együtthatót és az F-próba értékét növeli, akkor javítottunk a modellen. Ha a modell illeszkedése javul, akkor a becslés szórása is csökkenni fog.

A REGRESSZIÓSZÁMÍTÁS menüjéből a TÖBBVÁLTOZÓS LINEÁRIS REGRESSZIÓ A VÁLTOZÓK TRANSZFORMÁCIÓJÁVAL ágat választjuk. A program, annak érdekében, hogy több regressziós modell egymást követő vizsgálatát lehetővé tegye ugyanarra az adathalmazra, eleve feltételezi, hogy adataink lemezen vannak, és a bejelentkezés után mindjárt az adatfile nevét kérdezi. A név megadása után a program bekéri az adatlemezt. Az adathalmaz neve jelen esetben AMMON.

A név megadása után a program beolvassa az adatokat, kiírja a talált adatcsoportok számát (21) és a független változók számát (3).

A programban a független változók jele x_0, x_1, x_2. A függő változó jele Y. A független változókból legfeljebb 23 aritmetikai kifejezést lehet képezni. Ezek jele X_a, X_b, .:., X_W.

A kívánt adatcsoportok kihagyását úgy oldja meg a program, hogy az adatpontok közül csak azokra illeszti a függvényt, amelyekre egy, a felhasználó által definiálható logikai kifejezés igaz (azaz értéke nem Ø). Az Y értéke is transzformálható, de erre példánkban nincs szükség. Az eredmények a képernyőn automatikusan, külön kérés nélkül megjelennek.

Esetünkben a megfelelő transzformációk:

$$X_a = X_0$$

$$X_b = X_1$$

$$X_c = X_1 X_1 .$$

A 21., figyelmen kívül hagyandó adatot az különbözteti meg a többitől, hogy ez az egyetlen, amelynél az x_0 értéke (az oxidáló levegőáram mennyisége) 70. Ezért a megfelelő logikai kifejezés: $x0 \neq 70$.

Az Y transzformációjául (és annak inverzéül is) önmagát adjuk meg. Ezzel minden szükséges információt megadtunk a programnak. Megjelenik a képernyőn egy OK (I/N) üzenet. Ha az eddigiekben hibát követtünk el valahol, most lehetőségünk van javítani. Ha igennel válaszolunk, a program megkérdezi az output file nevét, és a számolás elindul. Első lépése az adatválogatás. Ha helyesen írtuk fel a logikai kifejezést, éppen azokat a pontokat találja meg, amelyekre a függvényt illeszteni akartuk, esetünkben az első 20 pontot. A program a képernyőre kiírja a talált pontok számát, a program és az adatfile nevét, és a regressziós függvényt a becsült konstansokkal:

$$Y = -14{,}5145 + 0{,}77375 X_0 - 2{,}321 X_1 + 0{,}07625 X_1 X_1$$

A regressziós táblázat esetünkben a következőképpen alakul:

Négyzetösszeg		Szab. fok	Négyzetösszeg / szab. fok
Regressziós	1955,5	3	651,8
Maradék	107,1	16	6,7
Teljes	2062,5	19	
F-érték			97,4126
Determinációs koefficiens			0,9481
Korrelációs koefficiens			0,9737
A becslés standard deviációja			2,5868

KÉR PONTONKÉNTI ELLENŐRZÉST? (I/N)

Természetesen kérünk. Válaszképpen valamennyi mérési pont sorszáma, mért és számított Y értéke és a kettő eltérése jelenik meg a képernyőn:

Sorszám	Mért	Számított	Eltérés
1	42	40	2
2	37	40	−3
3	37	33	4
4	28	22	6
5	18	19	−1
6	18	20	−2
7	19	22	−3
8	20	22	−2
9	15	17	−2
10	14	13	1
11	14	13	1
12	13	13	0
13	11	13	−2
14	12	14	−2
15	8	7	1
16	7	7	0
17	8	8	0
18	8	8	0
19	9	8	1
20	15	13	2

Átlagos eltérés = 2,54

Összehasonlítva az eredeti többváltozós lineáris regresszióval kapott eredményekkel (6.3. alfejezet), kitűnik, hogy bár ugyanannyi együtthatót tartalmazó egyenlettel közelítettük az összefüggést, az F-érték mégis 58-ról 97-re nőtt, azaz jelentős javulást értünk el. Hasonlóképpen nőtt a determinációs és korrelációs együtthatók értéke is: a becslő függvény a vizsgált független változó tartományban az Y változásának közel 95%-át visszavezeti az X változók adott függvény szerinti változására.

A program ezután újabb számítás lehetőségét ajánlja fel. Most meg lehet változtatni a függvény alakját, a pontok kihagyására felírt logikai kifejezést, stb. Ezt az eljárást addig érdemes ismételni, amíg az F-érték javítható. (Gondoljuk meg ugyanis a következőket: Ha a pontok közül egyet sem hagyunk el, akkor a teljes négyzetösszeg állandó. Ezt bontjuk fel két független négyzetösszegre, Q_1-re és Q_2-re. Minden kiírt statisztikai mérőszám ebből a két négyzetösszegből származtatható. Mivel a kettő összege konstans, az egyik csak úgy nőhet, ha a másik egyidejűleg csökken, így hányadosuk a legérzékenyebb mutatója minden változásnak. Másrészt ha F-értéke nő, akkor a determinációs és korrelációs együtthatók értéke is nő, a becslés szórása pedig csökken. Ezek nem független információk, így csak akkor kell értékükre újra figyelni, ha a regressziós számításba bevont pontok körét megváltoztatjuk.)

EGYVÁLTOZÓS POLINOM ILLESZTÉSE

A változók listája:

N — mérési pontok száma,
M — polinom fokszáma,
X(I) — független változó vektora,
Y(I) — függő változó vektora,
OX, OY — függő és független változó mátrixbeli oszlopszáma (lemezről történő adatbeolvasásnál használjuk),
Q — a polinom együtthatói
$(Y=Q(0)+Q(1)*X+Q(2)*X^2+\ldots+Q(M)*X^M)$,

S5 — determinációs együttható,
S6 — korrelációs együttható,
S7 — a becslés szórása.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—250:	A program neve.
260—720:	Adatbevitel.
730:	A polinom fokszámának megadása.
740—750:	A polinom fokszámának ellenőrzése.
760—1000:	A regressziós függvény együtthatóinak meghatározása.
1010—1030:	Az eredmény kiírási helyének kijelölése.
1040—1170:	Az együtthatók kiírása.
1180—1560:	A regressziós táblázat számítása.
1570—1730:	A mérési pontok ellenőrzése.
1740—1910:	Pontok meghatározása a regressziós görbéből.
1920:	Visszatérés a FŐMENÜ c. programhoz.
1930—2340:	Adatbeolvasás lemezről.
2350—2460:	Lemezhiba-ellenőrzés.
2470—2620:	A program által nyújtott lehetőségek rövid ismertetése.
2630—2810:	Várakozás Igen-Nem válaszra.

```
730 INPUT"A POLINOM FOKSZAMA";M
740 IF M<>INT(M) OR M<1 THEN PRINT"A FOKSZAM POZITIV EGESZ!":GOTO 730
750 IF N-M-1<0 THEN PRINT "A FOKSZAM LEGFELJEBB";N-1;"LEHET!":GOTO 730
760 PRINT:PRINT"SZAMOLAS"
770 DIM A(2*M),Q(M,M+1),E(M+1)
780 REM*******************************
790 REM*                             *
800 REM* EGYUTTHATOK MEGHATAROZASA   *
810 REM*                             *
820 REM*******************************
830 HS=0:A(0)=N
840 FOR I=1 TO N
850 FOR J=1 TO 2*M:A(J)=A(J)+X(I)↑J:NEXT J
860 FOR J=0 TO M:E(J)=E(J)+Y(I)*X(I)↑J:NEXT J
870 E(M+1)=E(M+1)+Y(I)*Y(I):NEXT I
880 FOR I=0 TO M:FOR J=0 TO M:Q(I,J)=A(I+J):NEXT J:Q(I,M+1)=E(I):NEXT I
890 FOR S=0 TO M:IF Q(S,S)<>0 THEN 950
900 IF S=M THEN 940
910 FOR T=S+1 TO M:IF Q(T,S)=0 THEN 930
920 FOR J=S TO M+1:B=Q(S,J):Q(S,J)=Q(T,J):Q(T,J)=B:NEXT J:T=S+1
930 NEXT T
940 IF Q(S,S)=0 THEN PRINT "NINCS EGYERTELMU MEGOLDAS!":HS=1:S=M:GOTO 990
950 C=1/Q(S,S):FOR J=S+1 TO M+1:Q(S,J)=C*Q(S,J):NEXT J
960 FOR T=0 TO M:IF T=S THEN 980
970 C=-Q(T,S):FOR J=S+1 TO M+1:Q(T,J)=Q(T,J)+C*Q(S,J):NEXT J
980 NEXT T
990 NEXT S
1000 IF HS<>0 THEN 1890
```

TÖBBVÁLTOZÓS LINEÁRIS REGRESSZIÓ A VÁLTOZÓK TRANSZFORMÁCIÓJÁVAL

Fontosabb változók listája:

N$	— az input adatfile neve,
N0	— az input adatmátrix sorai száma (mérési pontok száma),
M0	— az input adatmátrix oszlopainak száma (független változók száma),
BX(I, J)	— input adatmátrix,
X(N0, 27)	— a regressziószámításnál figyelembe veendő adatok mátrixa,
B(27)	— a közelítő függvény együtthatóinak vektora,
R(4, 3)	— regressziós adatok,
X0, ..., X9	— egy adatsor független változói,
Y	— egy adatsor függő változója,
A(27, 27)	— az együttható mátrix a Gauss—Jordan eljárásban.

Programmagyarázat:

100:	Programcsomag-szervezés.
110:	Szubrutin: felhasználói függvények létrehozása.
120:	Szubrutin: felhasználói függvények hívása.
130—150:	Szubrutin: x. felhasználói függvény keresése a memóriában.
160—170:	Szubrutin: felhasználói függvény cím változtatás a memóriában.
180—470:	Szubrutin: felhasználói függvények megadása.
180—190:	*x*. *Y* kiírása és a függvény képletének inputja.
200—460:	A függvény „tokenizálása".
470:	Az átalakított képlet elhelyezése az ideiglenes tárolóterületen.
480—530:	Szétugrató tábla.
540—1180:	Szubrutin: *m* változós lineáris regressziószámítás *n* adatsor alapján. Belépéskor az $x(n, m+1)$ mátrix tartalmazza az adatsorokat. Kilépéskor a $b(m+1)$ vektor tartalmazza a közelítő függvény együtthatóit, $r(4, 3)$ a regressziós adatokat.
540—880:	A megoldandó lineáris egyenletrendszer mátrixának kialakítása.
890—1060:	A lineáris egyenletrendszer megoldása Gauss—Jordan módszerrel, szükség esetén oszlopcserékkel.
1070—1180:	A regresszió jellemző adatainak kiszámítása.
1190—1240:	Szubrutin: behelyettesítés a közelítő függvénybe.
1250—1640:	Szubrutin: az eredmények kiírása.
1650—1700:	Szubrutin: kiírás a képernyőre. + megszakítási kérés figyelés (megszakítás a COMMODORE gombbal lehetséges).
1710—2000:	Szubrutin: kerekítés. Az *x* szám *ee* egész jegyet és *dd* tizedesjegyet tartalmazó kiírási képe lesz kilépéskor az *x*$ változóban (hossza $ee+dd+2$ mindig).
2010—2120:	Indulási képernyő.
2130—2140:	Az adatfile nevének bekérése.
2150—2250:	A kiírási képek megadása.
2140—2300:	Adatmátrix beolvasás és megjelenítés a képernyőn.
2250—2390:	Használati utasítás.

2400—2530: A változók transzformálására szolgáló függvények és a figyelembe veendő adatsorokat leíró logikai kifejezés bekérése.
2540—2710: A megadott feltételnek eleget tevő pontok kiválogatása, megjelenítése a képernyőn és elhelyezése az $x(\ )$ mátrixban.
2720: Szubrutinhívások: regressziószámítás és az eredmények kiírása.
2730—2740: Újabb számítás? Igen válasz esetén vissza a felhasználói függvények megadásához.
2750: A program vége.
2760—3070: Az adatok bevitele.
3080—3220: Lemezhiba-ellenőrzés.

```
2250 PRINT"A PROGRAMBAN A FUGGETLEN VALTOZOK JELE":PRINT"X0";
2260 IFM0>1THENFORI=1TOM0-1:PRINT",X"CHR$(48+I);:NEXT
2270 PRINT".":PRINT"A FUGGO VALTOZO JELE Y."
2280 PRINT"A FUGGETLEN VALTOZOKBOL LEGFELJEBB 23"
2290 PRINT"ARITMETIKAI KIFEJEZEST LEHET KIALAKITANI";
2300 PRINT"ES AZ OSSZEFUGGEST EZEK LINEARIS KOMBI-"
2310 PRINT"NACIOJAVAL KOZELITENI. JELUK":PRINT"XA,XB,...,XW."
2320 PRINT"AZ ADATPONTOK KOZUL CSAK AZOKRA ILLESZ-"
2330 PRINT"TUNK FUGGVENYT, AMELYEKRE A MEGADOTT XX"
2340 PRINT"LOGIKAI KIFEJEZES IGAZ (VAGYIS, NEM 0)."
2350 PRINT"AZ Y ERTEK IS TRANSZFORMALHATO; AZ UJ"
2360 PRINT"ERTEK JELE XY.
2370 PRINT"A TRANSZFORMACIO INVERZ FUGGVENYET"
2380 PRINT"IS MEG KELL ADNI, JELE XZ."
2390 PRINT"AZ EREDMENYEK KI IS NYOMTATHATOK."
2400 XX$="-1":XY$="Y":XZ$="XY"
2410 PRINT:PRINT"■ADJA MEG A KIFEJEZESEKET! (VEGE=URES)"
2420 CI=49152:J=0
2430 X$=X$(J):GOSUB180:X$(J)=X$:IF(X$="")THENM=J:GOTO2460
2440 IFJ=22THENM=23:GOTO2460
2450 J=J+1:GOTO2430
2460 J=23:X$=XX$:PRINT"■ADJA MEG AZ XX LOGIKAI KIFEJEZEST!":GOSUB180:XX$=X$
2470 PRINT:PRINT"■ADJA MEG AZ Y TRANSZFOMACIOJAT!":X$=XY$:J=24:GOSUB180:XY$=X$
2480 X$=XZ$
2490 PRINT:PRINT"■ADJA MEG AZ Y TRANSZFORMACIO INVERZET!":J=25:GOSUB180:XZ$=X$
2500 L$="N":INPUT"OK (I/N) ? N■■■";L$:IFL$<>"I"THEN2410
2510 INPUT"KEPERNYO(3) VAGY NYOMTATO(4)";A
2520 IF A<>3 AND A<>4 THEN 2510
2530 OPEN2,4:IN=2
2540 T$="ADATVALOGATAS":GOSUB520:GOSUB520:N=0:FORI=1TON0
2550 ONM0GOTO2650,2640,2630,2620,2610,2600,2590,2580,2570,2560
2560 X9=BX(I,9)
2570 X8=BX(I,8)
2580 X7=BX(I,7)
2590 X6=BX(I,6)
2600 X5=BX(I,5)
2610 X4=BX(I,4)
2620 X3=BX(I,3)
2630 X2=BX(I,2)
2640 X1=BX(I,1)
2650 X0=BX(I,0)
2660 Y=BX(I,M0)
2670 J=23:GOSUB120:IFNOT(X)THEN2710
2680 N=N+1:X(N,0)=I:FORJ=0TOM-1:GOSUB120:X(N,J+1)=X:NEXTJ:J=24:GOSUB120
2690 X(N,M+1)=X
2700 T$=STR$(X(N,0))+":":FORJ=0TOM0:T$=T$+STR$(BX(I,J)):NEXTJ:GOSUB520
2710 NEXTI:GOSUB520:T$=STR$(N)+" DB PONTOT TALALTUNK.":GOSUB520:GOSUB520
2720 T$="REGRESSZIO SZAMITAS":GOSUB520:GOSUB520:GOSUB490:GOSUB510:CLOSE2
2730 PRINT:PRINT"UJABB SZAMITAS AZ EDDIGI ADATOKKAL  (U)"
2740 PRINT:PRINT"UJABB SZAMITAS MAS ADATOKKAL        (M)"
2750 PRINT:PRINT"VEGE                                (V)"
```

6.5. Korrelációs mátrix számítása

A korrelációs együttható fogalmával a regressziós egyenes statisztikai vizsgálata kapcsán már megismerkedtünk. Az ott elmondottakat két valószínűségi változó összefüggésének vizsgálatára alkalmaztuk. Ha többváltozós problémát vizsgálunk (pl. a többváltozós lineáris regressziós egyenlet illesztésekor), akkor ez az összefüggésvizsgálat minden lehetséges változópárral kapcsolatban elvégezhető. Ha 3 független változót tartalmaz egyenletünk, akkor vizsgálható az X_1-X_2, az X_2-X_3, az X_1-X_3, az X_2-X_3, az $Y-X_1$, $Y-X_2$ és az $Y-X_3$ összefüggése, pontosabban együttváltozása, korrelációja. Ha a változókra vonatkozó mérési adatokat táblázatos formába (mátrix alakba) rendezzük, ebből a táblázatból a lehetséges változópárosításoknak megfelelő korrelációs együtthatók ugyancsak táblázatos alakban egyszerű vektoralgebrai műveletekkel előállíthatók. Az így kapott táblázat neve: *korrelációs mátrix.*

Emlékezzünk rá: a korrelációs együttható tulajdonképpen a két változó kovarianciájának és a két változó szórásai szorzatának hányadosa. (Vektoralgebrai értelemben n mérési pont esetén a korrelációs együttható éppen a két n dimenziós vektor által bezárt szög koszinusza.) Egyetlen változó saját magával vett kovarianciája a variancia, azaz a szórásnégyzet. Így a korrelációt egyetlen változóra kiszámítva mind a számlálóban, mind a nevezőben X ugyanaz a variancia áll. Ezért bármely változó önmagával vett korrelációja 1. (Vö.: bármely vektor önmagával párhuzamos, így bezárt szögük 0°.)

A korrelációs mátrix szimmetrikus. Ez nyilvánvaló, ha meggondoljuk, hogy az X_1-X_2 pár korrelációja nem függhet a változók sorrendjének megválasztásától. Vezessük be Y helyett az általánosság kedvéért az X_{n+1} jelölést. Ezekkel a jelölésekkel azt mondhatjuk, hogy a korrelációs mátrix olyan szimmetrikus $(n+1)\cdot(n+1)$-es mátrix,

— amelynek i-edik sorában a j. oszlopban az X_i és X_j változók korrelációs koefficiense áll;
— a mátrix főátlójában csupa 1-es áll;
— a mátrix elemei legfeljebb 1 értékűek.

Miért fontos számunkra a többváltozós regressziószámításban a korrelációs mátrix ismerete? Tulajdonképpen azért, mert ezzel eldönthetjük, hogy mely változókat kell a modellben figyelembe venni, ill. mi a változók közötti „erősorrend" a függő változóra gyakorolt befolyás tekintetében. A modell „javításakor" a gyengébb hatású változókat megkísérelhetjük elhagyni, esetleg a korrelációszámítás szerint erősebb hatású változók kifejezéseivel helyettesíteni. A korrelációs mátrix alakja tehát a következő:

$$\begin{matrix}
1 & r_{x_1x_2} & r_{x_1x_3} \dots & r_{x_1x_n} & r_{x_1x_{n+1}} \\
r_{x_2x_1} & 1 & r_{x_2x_3} & r_{x_2x_n} & r_{x_2x_{n+1}} \\
\cdot & & 1 \dots & & \cdot \\
\cdot & & \dots & & \cdot \\
\cdot & & \dots 1 & & \cdot \\
r_{x_{n+1}x_1} & r_{x_{n+1}x_2} & r_{x_{n+1}x_3} & r_{x_{n+1}x_n} & 1
\end{matrix}$$

Az elmondottak megvilágítására tekintsünk ismét egy példát! Számítsuk ki a 6.2. al-

fejezetben ismertetett salétromsavgyári véggáztisztító üzem adataira a korrelációs mátrixot. Adatainkat az AMMON file-ból vesszük, amelyet mágneslemezen tároltunk. Először behívjuk a REGRESSZIÓSZÁMÍTÁS menüjéből a KORRELÁCIÓS MÁTRIX SZÁMÍTÁSA alprogramot. Erről kérhetünk ismertetést vagy anélkül is továbbmehetünk. Mivel adataink az AMMON file-ban vannak, értelemszerűen válaszolunk I-t az „adatok lemezen vannak vagy billentyűzetről akarjuk bevinni (I/N)" kérdésre, megadjuk az adatfile nevét és betesszük az adatlemezt. A program az adatokat beolvassa, majd megkérdezi, hogy kívánjuk-e az adatokat képernyőn megjeleníteni. Ha nemmel válaszolunk, vagy ezen az ágon végighaladva már megtekintettük adatainkat, a program megkérdezi, hová írja az eredményeket. Az eredményül kapott korrelációs mátrix a következő lesz:

1	0,78185	0,50014	0,91966
0,78185	1	0,39093	0,87550
0,50014	0,39093	1	0,39982
0,91966	0,87550	0,39982	1

A mátrixból látható, hogy az Y-nal a leggyengébben korrelált változó X_3 volt. Ezért hagytuk ki az előző fejezetben éppen ezt a változót a modellből. Választásunkat a korrelációszámítás alátámasztja.

KORRELÁCIÓS MÁTRIX SZÁMÍTÁSA

A változók listája:

N — mérési pontok száma (max. 100),
M — független változók száma (max. 10),
X(I, J) — független változók mátrixa (K=1-től N-ig, L=1-től M-ig megy),
R(I, J) — a korrelációs koefficiensek mátrixa,
N$ — az adatfile neve.

Programmagyarázat:

100: Programcsomag-szervezés.
110—230: A program neve.
240—780: Adatbevitel.
790—1020: A korrelációs mátrix számítása.
1030—1050: Az eredmény kiírási helyének kijelölése.
1060—1340: Az eredmények kiírása.
1350: Visszatérés a FŐMENÜ c. programhoz.
1360—1690: Adatbeolvasás lemezről.
1700—1810: Lemezhiba-ellenőrzés.
1820—1980: A program által nyújtott lehetőségek rövid ismertetése.
1990—2170: Várakozás Igen-Nem válaszra.

```
790 REM******************************
800 REM*                            *
810 REM* KORRELACIOS MATRIX         *
820 REM*                            *
830 REM******************************
840 K=0
850 K=K+1
860 FOR I=1 TO R
870 FORJ=1 TO R
880 R(I,J)=R(I,J)+X(K,I)*X(K,J)
890 NEXT J
900 A(I)=A(I)+X(K,I)
910 B(I)=B(I)+X(K,I)↑2
920 NEXT I
930 IF K<N THEN 850
940 FOR I=1 TO R
950 S=N*B(I)-A(I)↑2
960 FOR J=I TO R
970 R(I,J)=(N*R(I,J)-A(I)*A(J))/SQR(S*(N*B(J)-A(J)↑2))
980 R(J,I)=R(I,J)
990 NEXT J
1000 R(I,I)=1
1010 NEXT I
1020 FOR I=1 TO R:FOR J=1 TO R: R(I,J)=INT(R(I,J)*1E5+0.5)/1E5:NEXT J:NEXT I
```

7. AZ ADATKEZELŐ PROGRAM

Az előző fejezetekben ismertetett statisztikai programokkal általában igen sok adatot dolgozunk fel egyszerre. Nagyobb adathalmazok feldolgozásánál mindenképpen szükség van háttértárolóra. Így elkerülhetjük az adatok esetenkénti újragépelését. Ebben a fejezetben adatkezelő programunk telepítését és működését ismertetjük.

Egy adatkezelő programmal szemben általában a következő elvárásaink vannak: a program tetszőleges adatmátrixot tudjon beolvasni a billentyűzetről, és azt tárolni tudja lemezen. A program tegye lehetővé a hibás adatok javítását. Adjon lehetőséget az adatállomány bővítésére, ill. a fölöslegessé vált adatok törlésére az adathalmazból. Fontos funkciója továbbá az adatok megjelenítése, kiírása képernyőre vagy kinyomtatása papírra.

Nézzük meg, hogyan működik adatkezelő programunk! A főmenüből válasszuk az ADATFILE MŰVELETEK című programot! A program az előző programoknál már megszokott KÉR ISMERTETÉST kérdéssel jelentkezik be. Ha igent válaszolunk, akkor megjelenik a képernyőn egy rövid ismertetés a program működéséről. Ha nem kérünk ismertetést, akkor az adatkezelő program főmenüje jelenik meg a képernyőn. Ez a következő lehetőségek közül enged választani:

ÚJ ADATHALMAZT AKAR LÉTREHOZNI (U)
RÉGI ADATHALMAZZAL DOLGOZIK (R)
VÉGE (V)

Nézzük soronként ezeket a lehetőségeket!

Ha még nincs kész az adathalmazunk, hanem most szeretnénk létrehozni, akkor természetesen az U billentyűt kell lenyomnunk. Ennek hatására a program megkérdezi a létrehozandó adathalmaz nevét és a független változók számát. A független változók száma az adatmátrix oszlopainak számánál eggyel kevesebb, az utolsó oszlopba a függő változót helyezi el a program. Az adatmátrix sorainak száma az adatcsoportok számával lesz egyenlő. (Figyeljünk az elnevezésre. Az adathalmaz adatcsoportokból áll, egy-egy adatcsoport összefüggő (egy méréshez tartozó) adatokat tartalmaz. Annyi adatcsoportból áll tehát az adathalmaz, ahány független mérési pont adatait tartalmazza.)

Ezután a következő almenü jelenik meg a képernyőn:

ÚJ ADATCSOPORT MEGADÁSA (U)
RÉGI ADATCSOPORT KIÍRÁSA (K)
RÉGI ADATCSOPORT MÓDOSÍTÁSA (M)
RÉGI ADATCSOPORT TÖRLÉSE (T)
ADATMÁTRIX KIÍRÁSA (A)
FELVÉTEL (F)
VÉGE (V)

Újabb U választásra a program bekéri az első adatcsoport független és függő változóinak értékeit. Az értékek begépelése után felteszi a program a kérdést, JÓK AZ ADATOK? Igenlő válasz esetén visszatér az almenü elejére, ahol folytathatjuk ugyanazon adathalmaz következő adatcsoportjának megadását. Ha az adatok gépelése közben hibát vétettünk, akkor lehetőségünk van a hibás adatok javítására. Ezzel a módszerrel tehát a kívánt adathalmazt létrehozhatjuk.

Az adathalmaz létrehozása után azt nyilván szeretnénk lemezen is megőrizni, hisz éppen az a célunk, hogy ne kelljen minden egyes feldolgozáshoz az adatokat újra begépelnünk. Az adathalmaz lemezre vételéhez az almenüből a FELVÉTEL ágat kell választani. Az F billentyű lenyomására a program bekéri az adathalmaz nevét, ami lehet az előzőekben megadott, ill. itt még módunk van más nevet választani.

Megjegyezzük, hogy az adatokat célszerű külön lemezen tárolni. Lemezkezelés közben számtalan hiba léphet fel, amire egy adatkezelő programot fel kell készíteni. Elképzelhető például, hogy formálatlan lemezt helyezünk be az egységbe adatlemezként, vagy a lemezünkön volt már olyan nevű adathalmaz, amit most megadtunk. Az ilyen típusú hibáknál a lemezegység piros lámpája villogni kezd. A lemezhibák programbeli kezelését a hibacsatorna segítségével végezzük. A hibacsatornán keresztül kiolvashatjuk a hiba számát és a hibaüzenetet, így a program működését ettől függően folytathatjuk.

Sikeres felvétel után a program az almenüre tér vissza, ahonnan — ha munkánk ezzel véget ért — a V billentyű lenyomásával léphetünk ki.

Nézzük ezután az almenü következő pontját A RÉGI ADATCSOPORT KIÍRÁSA-t! Erre a K betű lenyomásával van lehetőségünk. Az előzőekben bevitt adatcsoport, ill. adatcsoportok már meglevőnek, azaz réginek tekinthetők. Ha egy régi adatcsoport kiírásánál rájövünk, hogy az ott szereplő értékek nem helyesek (hibásan írtuk be azokat vagy időközben megváltoztak), akkor az M billentyű lenyomásával a RÉGI ADATCSOPORT MÓDOSÍTÁSA programrészt aktivizálhatjuk. Ez a programrész az ADATCSOPORT SZÁMA? kérdéssel jelentkezik be. Itt kell megadnunk annak az adatcsoportnak a sorszámát, amit módosítani szeretnénk. Ha nem emlékszünk rá, hogy hányadik adatcsoportról van szó, akkor az adathalmaz kiírásával ezt megtudhatjuk. A javítandó adatcsoport számának begépelése után az értékek megjelennek a képernyőn, mégpedig úgy, hogy a kurzor villogása jelzi az éppen javítható értéket. Így a javítást, ill. módosítást igen könnyen elvégezhetjük. Előfordulhat az is, hogy egy egész adatcsoportot törölni szeretnénk. Ezt az almenü ADATCSOPORT TÖRLÉSE programággal a T billentyű lenyomásával kezdeményezhetjük.

Nézzük végül az almenü utolsó, de igen fontos lehetőségét, a teljes adathalmaz kiírását! Az A billentyű lenyomásával az ADATMÁTRIX KIÍRÁSA programrész lép működésbe. A programrész a KÉPERNYŐ (3), PRINTER (4)? kérdéssel jelentkezik. Itt választhatjuk tehát ki az output perifériát. Ha csak azért íratjuk ki az adathalmazt, hogy szemünkkel átfutva gyorsan ellenőrizzünk valamit, akkor elegendő a képernyőre kiírás. Ha szeretnénk listán is megőrizni az adatokat, akkor a nyomtatót kell választanunk.

A főmenü második ágának választása a RÉGI ADATHALMAZZAL DOLGOZIK ág első tevékenysége egy lemezen már meglevő adathalmaz beolvasását teszi lehetővé. Az R billentyű lenyomása után az ADATHALMAZ NEVE kérdés jelenik meg a képernyőn. Itt kell megadnunk tehát annak az adathalmaznak a nevét, amivel a továbbiakban dolgozni szeretnénk. Ezután a TEGYE BE AZ ADATLEMEZT ÉS NYOMJON MEG EGY BILLENTYŰT! parancs kerül kiírásra. A szóköz billentyű

lenyomásának hatására megkezdődik az adott nevű adathalmaz beolvasása a lemezről a memóriába. Ezt a műveletet addig kell folytatni, amíg az utolsó adatot is beolvastuk. Az utolsó adat elérését az ST rendszerváltozó értéke jelzi, amelynek tartalmát a lemezegység operációs rendszere minden beolvasási lépésben módosítja. A file utolsó adatának beolvasásakor az ST rendszerváltozó értéke 64. Így a beolvasó ciklust az ST=64 vizsgálattal vezérelhetjük.

Ezek után a már beolvasott adathalmazzal az almenüben szereplő, az előzőekben bemutatott műveleteket végezhetjük. Az adatkezelő programból a teljes kilépést a V billentyű lenyomásával, a VÉGE funkció aktivizálásával érhetjük el.

ADATKEZELŐ PROGRAM

A változók listája:

N — mérési pontok száma,
M — független változók száma,
X(I, J) — a függő és független változók mátrixa ($I=1$-től N-ig, $J=1$-től $M+1$-ig megy),
N$ — az adatfile neve.

Programmagyarázat:

100:	Programcsomag-szervezés.
110—230:	A program neve.
240—470:	Főmenü.
480—500:	Független változók számának megadása.
510—540:	Alprogramok hívása.
550—760:	Adatbeolvasás.
770—960:	Adatfelvétel.
970—1270:	Almenük felsorolása, vezérlésátadás a különböző almenüknek.
1280—1500:	Új adatcsoport megadása.
1610—1790:	Régi adatcsoport kiírása.
1800—1860:	Régi adatcsoport módosítása.
1870—2040:	Régi adatcsoport törlése.
2050—2190:	Lemezhiba ellenőrzése.
2200—2360:	Adatcsoport kiírása.
2370—2640:	Adathalmaz kiírása.
2650—2880:	A program által nyújtott lehetőségek rövid ismertetése.
2890—3030:	Várakozás Igen-Nem válaszra.

```
100 POKE45,PEEK(174):POKE46,PEEK(175):CLR
110 PRINT CHR$(147)
120 PRINT"********************************"
130 PRINT"*                              *"
140 PRINT"*  ADATKEZELO PROGRAM          *"
150 PRINT"*                              *"
160 PRINT"********************************"
170 PRINT
180 PRINT"********************************"
190 PRINT"*                              *"
200 PRINT"* KESZITETTE:BALINT AGNES      *"
210 PRINT"*            TATRAI FERENC     *"
220 PRINT"*                              *"
230 PRINT"********************************"
240 PRINT
250 PRINT"KER ISMERTETEST";:A$="N":GOSUB 3010
260 IF A$="I" THEN GOSUB 2650
270 VI$="████████████████████████████████"
280 MX=10:NX=200
290 DIM X(NX,MX+1)
300 PRINT CHR$(147);
310 REM*******************************
320 REM*                             *
330 REM*  FOMENU:                    *
340 REM*  UJ ADATHALMAZ.LETREHOZASA  *
350 REM*  MUVELETEK REGI ADATHALMAZZAL*
360 REM*                             *
370 REM*******************************
380 PRINT"UJ ADATHALMAZT AKAR LETREHOZNI (U)"
390 PRINT"REGI ADATHALMAZZAL DOLGOZIK    (R)"
400 IF N$<>"" THEN PRINT"TOVABBI JAVITASOK                (T)"
410 PRINT"VEGE                           (V)"
420 PRINT
430 INPUT"VALASZTASA";A$:IF A$<>"U" AND A$<>"R" AND A$<>"V" AND A$<>"T" THEN 430
440 IFA$="T" AND N$="" THEN 430
450 IF A$="T" THEN GOSUB 970:GOTO 300
460 IF A$="V" THEN POKE8*256,0:POKE44,8:POKE45,1:POKE46,30:RUN
470 IFA$<>"U"THEN510
480 INPUT"FUGGETLEN VALTOZOK SZAMA:";M:N=0
490 IF M>=0 AND M<=MX THEN 530
500 PRINT"HIBAS VALTOZOSZAM":GOTO 480
510 GOSUB 2400:IF N$="" THEN 300
520 GOSUB 600:IF HS>0 THEN 300
530 GOSUB 970
540 GOTO 300
550 REM*******************************
560 REM*                             *
570 REM*  ADATBEOLVASAS              *
580 REM*                             *
590 REM*******************************
600 IF N$="" THEN RETURN
610 GOSUB 2920
620 PRINT"ADATBEOLVASAS"
630 OPEN15,8,15:OPEN8,8,8,N$
640 GOSUB 2100:IFHS<>0THEN N$="":GOTO 750
650 INPUT#8,M
660 INPUT#8,N
670 IF N<1 THEN 750
680 IF M>MX OR N>NX THEN PRINT"TUL SOK ADAT":GOSUB 2930:RETURN
690 FOR I=1 TO N
700 :FOR J=1 TO M+1
710 ::INPUT#8,X(I,J)
720 ::GOSUB 2100:IF HS<> 0 THEN I=N:J=M+1
730 :NEXT J
740 NEXT I
750 CLOSE 8:CLOSE 15
760 RETURN
770 REM*******************************
780 REM*                             *
790 REM*  ADATFELVETEL               *
800 REM*                             *
810 REM*******************************
820 IF N$="" THEN RETURN
```

```
830 GOSUB 2920
840 OPEN15,8,15,"S:"+N$:OPEN8,8,8,N$+",S,W"
850 GOSUB 2100:IF HS<> 0 THEN 950
860 PRINT#8,M
870 PRINT#8,N
880 IF N=0 THEN 950
890 FOR I=1 TO N
900 :FOR J=1 TO M+1
910 ::PRINT#8,X(I,J)
920 ::GOSUB 2100: IF HS<>0 THEN I=N:J=M+1
930 :NEXT J
940 NEXT I
950 CLOSE 8:CLOSE 15
960 RETURN
970 REM******************************
980 REM*                            *
990 REM*  ALMENU:                   *
1000 REM*  ADATBEVITEL               *
1010 REM*  ADATMODOSITAS             *
1020 REM*  ADATTORLES                *
1030 REM*  ADATKIIRAS                *
1040 REM*                            *
1050 REM*****************************
1060 PRINT CHR$(147);"ADATHALMAZ :";N$
1070 PRINT
1080 PRINT"FUGGETLEN VALTOZOK SZAMA =";M
1090 PRINT"ADATCSOPORTOK SZAMA      =";N
1100 PRINT
1110 PRINT"UJ ADATCSOPORT MEGADASA      (U)"
1120 PRINT"REGI ADATCSOPORT KIIRASA     (K)"
1130 PRINT"REGI ADATCSOPORT MODOSITASA (M)"
1140 PRINT"ADATCSOPORT TORLESE          (T)"
1150 PRINT"ADATMATRIX KIIRASA           (A)"
1160 PRINT"FELVETEL                     (F)"
1170 PRINT"VEGE                         (V)"
1180 PRINT
1190 INPUT "VALASZTASA";B$
1200 IF B$="U" THEN GOSUB 1540:GOTO 1060
1210 IF B$="K" THEN GOSUB 1660:GOTO 1060
1220 IF B$="M" THEN GOSUB 1850:GOTO 1060
1230 IF B$="T" THEN GOSUB 1920:GOTO 1060
1240 IF B$="A" THEN GOSUB 2480:GOTO 1060
1250 IF B$="F" THEN GOSUB 2400:GOSUB 820:GOTO 1060
1260 IF B$="V" THEN GOTO 300
1270 GOTO 1060
1280 REM*****************************
1290 REM*                           *
1300 REM*  UJ ADATCSOPORT MEGADASA  *
1310 REM*                           *
1320 REM*****************************
1330 PRINT CHR$(147);
1340 PRINT"ADATHALMAZ :"N$
1350 PRINT
1360 PRINT"ADATCSOPORT:";N;"/";R
1370 PRINT
1380 IFM=0THEN1440
1390 FOR J=1 TO M
1400 PRINT "X(";MID$(STR$(J),2);")";TAB(7);"=";
1410 X(R,J)=0:INPUTX(R,J)
1420 NEXTJ
1430 PRINT
1440 PRINT"Y";TAB(7);"=";
1450 X(R,M+1)=0:INPUT X(R,M+1)
1460 PRINT
1470 PRINT "JOK AZ ADATOK";
1480 GOSUB 3010

1490 IF A$="I" THEN RETURN
1500 GOTO 1340
1510 REM*****************************
1520 REM*  UJ ADATCSOPORT HOZZAVETEL *
1530 REM*****************************
1540 IF N<NX THEN 1590
1550 PRINT
```

```
1560 PRINT"AZ ADATHALMAZ NEM BOVITHETO"
1570 GOSUB 2930
1580 RETURN
1590 N=N+1:R=N:GOSUB 1330
1600 RETURN
1610 REM*******************************
1620 REM*                             *
1630 REM*  REGI ADATCSOPORT KIIRASA   *
1640 REM*                             *
1650 REM*******************************
1660 GOSUB 1720:IF ER=1 THEN RETURN
1670 IF ER=1 THEN RETURN
1680 GOSUB 2250:RETURN
1690 REM*******************************
1700 REM*  ADATCSOPORT SZAM KERDES    *
1710 REM*******************************
1720 R=0
1730 IF N=0 THEN PRINT:PRINT "AZ ADATHALMAZ URES.":GOSUB 2930:ER=1:RETURN
1740 ER=0:INPUT"ADATCSOPORT SZAMA :";R
1750 IF R>0 AND R<=N THEN RETURN
1760 PRINT
1770 PRINT"NEM LETEZO ADATCSOPORT"
1780 GOSUB 2930:ER=1
1790 RETURN
1800 REM*******************************
1810 REM*                             *
1820 REM* REGI ADATCSOPORT MODOSITASA*
1830 REM*                             *
1840 REM*******************************
1850 GOSUB 1720:IF ER=1 THEN RETURN
1860 GOSUB 2250:GOSUB 1340:RETURN
1870 REM*******************************
1880 REM*                             *
1890 REM* REGI ADATCSOPORT TORLESE    *
1900 REM*                             *
1910 REM*******************************
1920 GOSUB 1720
1930 IF ER=1 THEN RETURN
1940 PRINT"BIZTOS, HOGY TOROLJUK";
1950 GOSUB 3010
1960 IF A$="N" THEN RETURN
1970 IF R=N THEN 2030
1980 FOR I=R TO N-1
1990 :FOR J=1 TO M+1
2000 ::X(I,J)=X(I+1,J)
2010 :NEXT J
2020 NEXT I
2030 N=N-1
2040 RETURN
2050 REM*******************************
2060 REM*                             *
2070 REM* LEMEZHIBA ELLENORZESE       *
2080 REM*                             *
2090 REM*******************************
2100 INPUT#15,HS,HN$,S,SZ
2110 IF HS=0 THEN RETURN
2120 PRINT
2130 PRINT"LEMEZHIBA!":PRINT
2140 PRINT HS;HN$;S;SZ
2150 PRINT
2160 GOSUB 2930:RETURN
2170 PRINT"NYOMJA MEG A MEGFELELO GOMBOT!"
2180 GETA$:IFA$=""THEN2180
2190 RETURN
2200 REM*******************************
2210 REM*                             *
2220 REM* ADATCSOPORT KIIRASA         *
2230 REM*                             *
2240 REM*******************************
2250 PRINTTCHR$(147);"ADATHALMAZ :"N$
2260 PRINT
2270 PRINT"ADATCSOPORT:";N;"/";R
2280 PRINT
2290 IFM=0 THEN2340
2300 FOR J=1 TO M
```

```
2310 PRINT "X(";MID$(STR$(J),2);")";TAB(7);"= ";X(R,J)
2320 NEXTJ
2330 PRINT
2340 PRINT"Y";TAB(7);"= ";X(R,M+1)
2350 IF B$<>"M" THEN GOSUB 2930
2360 RETURN
2370 REM******************************
2380 REM*  ADATFILE NEV BEKERES      *
2390 REM******************************
2400 X$="":PRINT"AZ ADATHALMAZ NEVE:? ";N$;LEFT$(VI$,LEN(N$)+2);
2410 INPUT X$:IF N$="" AND X$="" THEN 2400
2420 N$=X$:RETURN
2430 REM******************************
2440 REM*                            *
2450 REM* ADATHALMAZ KIIRASA         *
2460 REM*                            *
2470 REM******************************
2480 PRINT
2490 INPUT"KEPERNYO(3),PRINTER(4)";ID
2500 IF ID<>3 AND ID<>4 THEN 2490
2510 OPENID,ID:SN=6:IF ID=3 THEN PRINT CHR$(147);:SN=3
2520 PRINT#ID,"ADATMATRIX :";N$
2530 PRINT#ID:LN=2:IF N=0 THEN 2630
2540 FOR I=1 TO N
2550 PRINT#ID,RIGHT$("    "+STR$(I),3);:S=0
2560 FOR J=1 TO M+1
2570 PRINT#ID,RIGHT$("                "+STR$(X(I,J)),12);
2580 S=S+1:IF S=SN AND J<>M+1 THEN PRINT#ID:PRINT#ID,"    ";:S=0:LN=LN+1
2590 NEXT J
2600 PRINT#ID:LN=LN+1
2610 IF ID=3 AND LN+INT((M+1)/3)>20 AND I<>N THENGOSUB2930:LN=0:PRINT CHR$(147);
2620 NEXT I
2630 CLOSE ID:GOSUB 2930
2640 RETURN
2650 REM******************************
2660 REM*                            *
2670 REM*  ISMERTETES                *
2680 REM*                            *
2690 REM******************************
2700 PRINT CHR$(147)
2710 PRINT"ADATKEZELO PROGRAM"
2720 PRINT"------------------"
2730 PRINT"A PROGRAM AZ ALABBI FELADATOK VEGZESERE ALKALMAS:"
2740 PRINT"              UJ ADATHALMAZ LETREHOZASA"
2750 PRINT"              ADATHALMAZ FELVETELE LEMEZRE"
2760 PRINT"              REGI ADATHALMAZ BETOLTESE"
2770 PRINT"                 LEMEZROL"
2780 PRINT
2790 PRINT"              ADATHALMAZ MODOSITASA"
2800 PRINT"              ADATHALMAZ BOVITESE"
2810 PRINT"              ADATCSOPORT  TORLESE AZ ADAT"
2820 PRINT"                 HALMAZBOL"
2830 PRINT
2840 PRINT"              ADATHALMAZ KEPERNYORE IRASA"
2850 PRINT"              ADATHALMAZ NYOMTATORA IRASA"
2860 PRINT"              ADATCSOPORT KEPERNYORE IRASA"
2870 PRINT"              ADATCSOPORT NYOMTATORA IRASA"
2880 GOSUB 2930:RETURN
2890 REM**************************
2900 REM*     VARAKOZAS          *

2910 REM**************************
2920 PRINT:PRINT"TEGYE BE AZ ADATLEMEZT"
2930 PRINT:PRINT"NYOMJON MEG EGY BILLENTYUT!"
2940 FOR BQ=1 TO 16:GET X$:NEXT BQ
2950 GET X$:IF X$="" THEN 2950
2960 PRINT"]                                      ":PRINT"]";
2970 RETURN
2980 REM**************************
2990 REM* IGEN/NEM KERDES        *
3000 REM**************************
3010 INPUT " (I/N)";A$
3020 IF A$<>"I" AND A$<>"N" THEN 3010
3030 RETURN

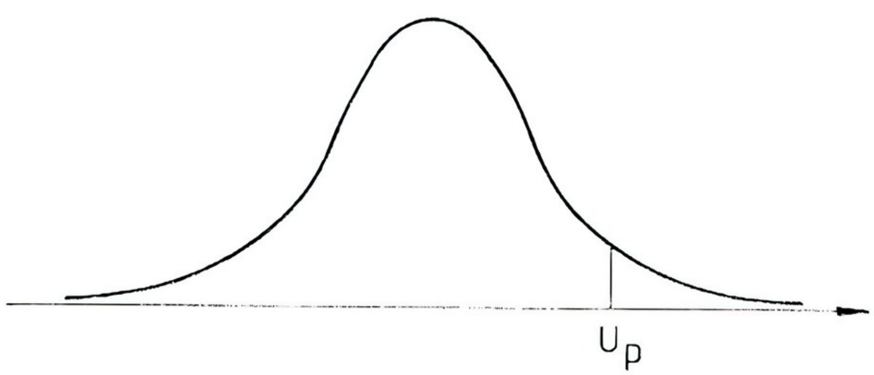
READY.
```

MELLÉKLET

1. táblázat

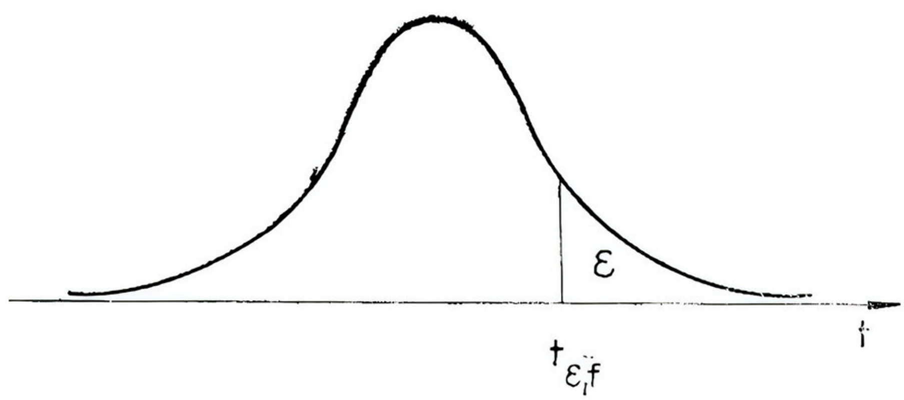
A normális eloszlás kritikus értékei (u_p)

(u_p)	,00	,01	,03	,05	,07	,09
0,0	,5000	,5040	,5120	,5199	,5279	,5359
,1	,5398	,5438	,5517	,5596	,5675	,5753
,2	,5793	,5832	,5910	,5987	,6064	,6141
,3	,6179	,6217	,6293	,6368	,6443	,6517
,4	,6554	,6591	,6664	,6736	,6808	,6879
,5	,6915	,6950	,7019	,7088	,7157	,7224
,6	,7257	,7291	,7357	,7422	,7486	,7549
,7	,7580	,7611	,7673	,7734	,7793	,7852
,8	,7881	,7910	,7967	,8023	,8078	,8133
,9	,8159	,8186	,8238	,8289	,8340	,8389
1,0	,8413	,8438	,8485	,8531	,8577	,8621
1,1	,8643	,8665	,8708	,8749	,8790	,8830
1,2	,8849	,8869	,8906	,8943	,8980	,9015
1,3	,9032	,9049	,9082	,9115	,9147	,9177
1,4	,9192	,9207	,9236	,9265	,9292	,9319
1,5	,9332	,9345	,9370	,9394	,9418	,9441
1,6	,9452	,9463	,9484	,9505	,9525	,9545
1,7	,9554	,9564	,9582	,9599	,9616	,9633
1,8	,9641	,9648	,9664	,9678	,9693	,9706
1,9	,9713	,9719	,9732	,9744	,9756	,9767
2,0	,9772	,9778	,9788	,9798	,9808	,9817
2,1	,9821	,9826	,9834	,9842	,9850	,9857
2,2	,9861	,9864	,9871	,9878	,9884	,9890
2,3	,9893	,9896	,9901	,9906	,9911	,9916
2,4	,9918	,9920	,9924	,9929	,9932	,9936
2,5	,9938	,9940	,9943	,9946	,9949	,9952
2,6	,9953	,9955	,9957	,9960	,9962	,9964
2,7	,9965	,9966	,9968	,9970	,9972	,9974
2,8	,9974	,9975	,9977	,9978	,9979	,9981
2,9	,9981	,9982	,9983	,9984	,9985	,9986
3,0	,9986	,9987	,9988	,9989	,9989	,9990
3,2	,9993	,9993	,9994	,9994	,9995	,9995
3,4	,9997	,9997	,9997	,9997	,9997	,9998
3,6	,9998	,9998	,9999	,9999	,9999	,9999



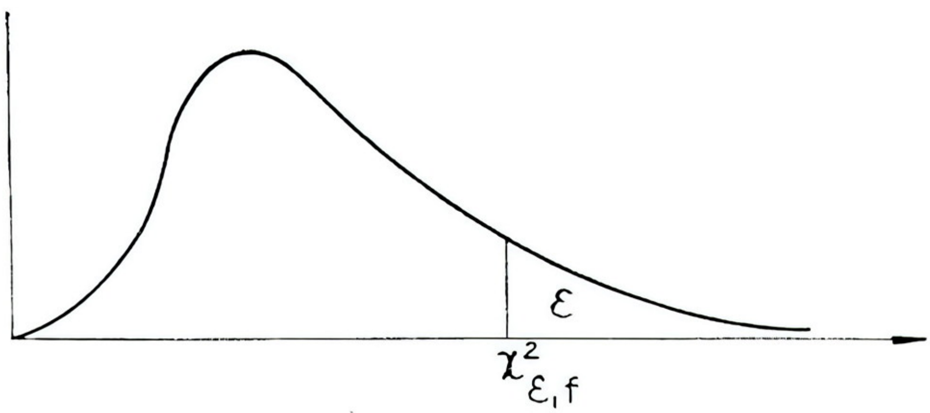
A Student eloszlás kritikus értékei ($t_{\varepsilon, f}$)

ε	,10	,05	,025	,01	,005	,001
f	$t_{\varepsilon, f}$					
1	3,078	6,314	12,706	31,621	63,657	318,310
2	1,886	2,920	4,303	6,965	9,925	22,327
3	1,638	2,353	3,182	4,541	5,841	10,215
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,785
8	1,397	1,860	2,306	2,896	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,812	2,228	2,764	3,169	4,144
15	1,341	1,753	2,131	2,602	2,947	3,733
20	1,325	1,725	2,086	2,528	2,845	3,552
30	1,310	1,697	2,042	2,457	2,750	3,385
40	1,303	1,684	2,021	2,423	2,704	3,307
60	1,296	1,671	2,000	2,390	2,660	3,232
120	1,289	1,658	1,980	2,358	2,617	3,160
∞	1,282	1,645	1,960	2,326	2,576	3,090



A χ^2 eloszlás kritikus értékei $\chi^2_{\varepsilon, f}$

ε	,995	,99	,975	,95	,50	,20	,10	,05	,025	,01	,005
f	$\chi^2_{\varepsilon, f}$										
1	0,000	0,0002	0,001	0,0039	0,45	1,64	2,71	3,84	5,02	6,63	7,88
2	0,010	0,020	0,051	0,103	1,39	3,22	4,61	5,99	7,38	9,21	10,60
3	0,072	0,115	0,216	0,352	2,37	4,64	6,25	7,81	9,35	11,34	12,84
4	0,207	0,30	0,484	0,71	3,36	5,99	7,78	9,49	11,14	13,28	14,86
5	0,412	0,55	0,831	1,15	4,35	7,29	9,24	11,07	12,83	15,09	16,75
6	0,676	0,87	1,24	1,64	5,35	8,56	10,64	12,59	14,45	16,81	18,55
7	0,989	1,24	1,69	2,17	6,35	9,80	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	7,34	11,03	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	3,33	8,34	12,24	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	9,34	13,44	15,99	18,31	20,48	23,21	25,19
15	4,60	5,23	6,26	7,26	14,34	19,31	22,31	25,00	27,49	30,58	32,80
20	7,43	8,26	9,59	10,85	19,34	25,04	28,41	31,41	34,17	37,57	40,00
25	10,52	11,52	13,12	14,61	24,34	30,68	34,38	37,65	40,65	44,31	46,93
30	13,79	14,95	16,79	18,49	29,34	36,25	40,26	43,77	46,98	50,89	53,67
40	20,71	22,16	24,43	26,51	39,34	47,27	51,81	55,76	59,34	63,69	66,77
50	27,99	29,71	32,36	34,76	49,33	58,16	63,17	67,50	71,41	76,15	79,49
60	35,53	37,48	40,48	43,19	59,33	68,97	74,40	79,08	83,30	88,38	91,95
70	43,28	45,44	48,76	51,74	69,33	79,71	85,53	90,53	95,02	100,43	104,2
80	51,17	53,54	57,15	60,39	79,33	90,41	96,58	101,88	106,63	112,33	116,3
90	59,20	61,75	65,65	69,13	89,33	101,05	107,57	113,15	118,14	124,12	128,3
100	67,33	70,06	74,22	77,93	99,33	111,67	118,50	124,34	129,56	135,81	140,2

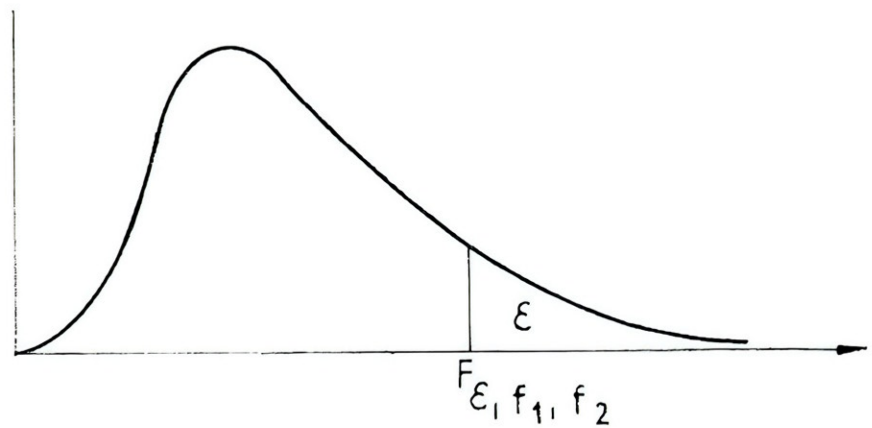


4. táblázat

A Fisher eloszlás kritikus értékei ($F_{\varepsilon, f_1 f_2}$)

$\varepsilon = 0{,}01$

f_1 / f_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30
1	4052,2	4999,5	5403,4	5624,6	5763,6	5859,0	5928,4	5981,1	6022,5	6055,8		6157,3	6208,7	6234,6	6260,6
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40		99,43	99,45	99,46	99,47
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23		26,87	26,69	26,60	26,50
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55		14,20	14,02	13,93	13,84
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05		9,72	9,55	9,47	9,38
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87		7,56	7,40	7,31	7,23
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62		6,31	6,16	6,07	5,99
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81		5,52	5,36	5,28	5,20
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26		4,96	4,81	4,73	4,65
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85		4,56	4,41	4,33	4,25
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80		3,52	3,37	3,29	3,21
20	8,19	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37		3,09	2,94	2,86	2,78
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13		2,85	2,70	2,62	2,54
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98		2,70	2,55	2,47	2,39
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80		2,52	2,37	2,29	2,20
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63		2,35	2,20	2,12	2,03
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47		2,19	2,03	1,95	1,86
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32		2,04	1,88	1,79	1,70



$\varepsilon = 0,05$

f_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30
f_2															
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91	245,95	248,01	249,05	250,10
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,84	2,77	2,74	2,70
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46

$\varepsilon = 0,025$

f_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30
f_2															
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28	968,63	976,71	984,87	993,10	997,25	1001,4
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45	39,46	39,46
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17	14,12	14,08
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,75	8,66	8,56	8,51	8,46
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,28	6,23
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,37	5,27	5,17	5,12	5,07
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,41	4,36
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,95	3,89
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,61	3,56
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,37	3,31
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,86	2,76	2,70	2,64
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,41	2,35
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,27	2,21
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,30	2,24	2,18
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,14	2,07
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	2,01	1,94
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,88	1,82
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,76	1,69
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,64	1,57